MODELING OF VOCAL FOLD VIBRATION

By

KWEI CHAN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1989

TO MY FAMILY

ACKNOWLEDGMENTS

I am deeply grateful to Professor D. G. Childers for supervising this dissertation. Its inception and ultimate completion would not have been possible without his constant guidance, encouragement, and support.

I also wish to thank Dr. G. Paul Moore for many stimulating discussions about speech research. My thanks are extended to Dr. F. J. Taylor, Dr. C. C. Formby, Dr. H. B. Rothman and Dr. J. R. Smith for serving on my supervisory committee. I thank all my colleagues in the Mind-Machine Interaction Research Center for help in many ways.

I especially wish to thank my parents for their love and support throughout the years. I also apologize to my wife, Mihchyn, and children, Daniel and Stephen, for their loss of my time.

TABLE OF CONTENTS

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of
the Requirements for the Degree of Doctor of Philosophy

MODELING OF VOCAL FOLD VIBRATION

By

Kwei Chan

August, 1989

Chairman: D. G. Childers
Major Department: Electrical Engineering

This research has resulted in the development of a unique kinematic model that characterizes the vibratory motion of the vocal folds in terms of a three-dimensional glottal configuration with spatially varying tissue properties. The model has been used to 1) establish relationships between the vibratory pattern of the vocal folds and several glottographic waveforms, 2) model the influence of voice disorders on the vibratory pattern of the vocal folds, and 3) study the behavior of glottographic waveforms associated with various vocal registers.

The purpose of the vibratory model was to describe aspects of the kinematics of vocal fold vibration by incorporating several concepts of vibratory behavior of the vocal folds. These concepts included 1) pre-phonatory configuration of the glottis, 2) maximum excursion profiles along the length and

depth of the glottis, and 3) the time lags in movement along the length and depth of the glottis.

To interpret frequently used glottographic signals in relation to the vibratory pattern of the vocal folds, algorithms were developed to express the three glottographic waveforms (i.e., electroglottogram, photoglottogram, and the inverse-filtered glottal waveform) in terms of the displacement of the glottis.

Factors contributing to abnormal vibratory patterns (i.e., tension imbalance of the vocal folds, irregular shape of the vocal folds, and abnormal tissue properties) were investigated and associated glottographic waveforms were simulated with the vibratory model. Two common vibratory patterns of abnormal vocal folds (e.g., incomplete glottal closure and perturbed vibration) were related to vocal fold physiology and features of the glottographic waveforms.

Finally, the basic differences in the vibratory mechanisms among three vocal registers (e.g., modal, vocal fry, and falsetto) were studied.

CHAPTER 1

INTRODUCTION

The human larynx is the organ of the voice. During the production of speech, the vocal folds produce controlled gradations of fundamental frequency, intensity, and voice quality. The vocal folds are also responsible for producing acoustic contrasts in phonetic segments. The mechanisms which control these changes and contrasts may be represented in terms of a quantitative model of the vocal folds.

The purpose of this dissertation was two-fold. The first goal was to develop a vibratory model of the vocal folds. The aim of this model was to simulate and achieve many varieties of phonation. The second goal was to investigate the relations between the vibratory pattern of vocal folds, and the glottographic waveforms, especially the electroglottogram (EGG). The extensive use of non-invasive glottography in speech research has prompted efforts to interpret the glottographic signals and improve the instrumentation of these techniques. This proposed research has increased our understanding of the relation between vocal fold configurations and the voice characteristics of a speaker. As a part of the second goal, we also sought to evaluate the

usefulness of glottography as a tool for detection of vocal fold dysfunction and as an aid to speech analysis.

In Chapter 1, applications of voice excitation research, summaries of other approaches to this research, anatomy and physiology of the vocal folds, and models of vocal fold vibration will be discussed. This dissertation presents a new vocal fold vibratory model as a 3-dimensional vocal fold configuration. The model was used to study the relations among the vibratory patterns of the vocal folds, several glottographic waveforms, and the speech signals. The proposed kinematic model of vocal fold vibration, which is detailed in Chapter 2, was able to simulate the glottal area, the lateral contact area, and glottal volume velocity with the same parameter set. This procedure is discussed in Chapter 3. Therefore, the measured glottographic waveforms can be compared and studied against their synthetic counterpart. The relation between the glottographic signals and the vocal fold vibratory patterns under various phonating conditions is established using the proposed vibratory model.

In Chapter 4, the vibratory patterns of abnormal vocal folds are studied. Factors contributing to abnormal vibratory patterns, such as imbalance of the vocal folds, irregular shape of the vocal folds, and abnormal tissue properties, are investigated and simulated with the proposed vibratory model. The use of the EGG for detection of abnormal vocal fold function is also discussed.

In Chapter 5, we investigate how the vibratory patterns of the vocal folds vary with the vocal registers. Specifically, the relations between three vocal registers (i.e., modal, vocal fry, and falsetto) and the parameters of the glottographic waveforms are established.

In Chapter 6, a summary of this research and suggestions for future work are given.

Voice Excitation Research and Its Applications

An understanding of voice-producing mechanisms can benefit speech synthesis and analysis, linguistics, and clinical practice, including detection, diagnosis, and treatment of vocal disorders. Let us consider these different disciplines and how knowledge about voice excitation can be applied.

For speech synthesis and analysis, researchers are interested in the laryngeal characteristics that contribute to the naturalness and the vocal quality of speech. Naturalness of a speech segment is a highly subjective attribute that may often suggest the impression of "human sounding" [Klatt, 1987]. The quality of speech is usually referred to as the total auditory impression the listener experiences upon hearing the speech of a speaker and this impression may include naturalness as a factor.

Users of speech synthesizers expect the product to be infused with character, personality, and identity. Subtle differences in laryngeal behavior are believed to be crucial

in differentiating male voices from female voices, coarse voices from light voices, or breathy voices from harsh voices. However, sublaryngeal and supralaryngeal features may contribute to voice quality as well. Some important acoustic features of the glottal excitation have been identified for synthesis of natural sounding speech and for creating voices with a specific quality. Researchers have recognized, for example, the importance of noise components to the perception of breathiness, and the importance of slight cycle-to-cycle pitch variations to the naturalness of speech. However, additional features need to be identified and rules for the dynamic control of these features are still quite primitive. One of the major objectives of this research was to study how complex vibratory patterns of the vocal fold affect human voice characteristics. From these findings we hoped to suggest ways to improve the simulation of the source of the voiced sounds.

Phoneticians often focus on laryngeal adjustments and timing in speech production (e.g., glottal opening control in consonant production [Abramson, 1979] and interaction between voicing and pitch). An example for the latter case is that the rising fundamental frequency near voice onset in consonant-vowel (CV) syllables perceptually contributes to the phonemic distinction between consonants /p/ and /b/ [Fujimura, 1979]. In this area, the dynamics of the gross adjustments of the mechanisms of the phonating apparatus are the focus of

the research. Flanagan, Rabiner, Christopher, Bock, and Shipp [1976] used electromyography (EMG) to monitor activities of the intrinsic laryngeal muscles to gain knowledge for proper control of their articulatory synthesizer. Lofqvist and Yoshioka [1980] used the combined observation techniques of EMG, fiberoptic filming, and photoglottography (PGG) of the larynx for study of the laryngeal activity in consonant clusters.

There is a growing body of studies showing correlations between acoustic measures of the voice and pathological states of the larynx [Davis, 1979; Gauffin, Sundberg, and Hammarberg, 1980; Kasuya, Ogawa, Mashima, and Ebihara, 1986]. The fact that diseases of the larynx are often accompanied by changes in the speech of patients prompted a search for acoustic cues that might be useful for the detection of laryngeal diseases. However, the analysis performed on speech signals does not lend itself to clinical use, partly as a result of the limited size of patient data base and partly because the supraglottal structure may often mask some of the important acoustic attributes of a laryngeal pathology.

Researchers believe that the vibratory pattern of the vocal folds is a major factor relating laryngeal function to the sound produced. Therefore, detection of a laryngeal malfunction by acoustic analysis requires two stages of research. In the first stage, it is essential to understand the various normal and pathological modes of vibration and

their relation to physical states of the larynx. In the second stage, the relation between the pathological vibratory patterns of the vocal folds and various measured signals needs to be established.

Glottographic signals (i.e., EGG and PGG signals) are widely used for assessment of laryngeal function [Childers, Smith, and Moore, 1984; Rosenberg, 1983; Fourcin, 1976] and for speech analysis [Krishnamurthy and Childers, 1986] because of their abilities to monitor the vocal fold activity. In the search for a laryngeal counterpart to electrocardiography, we need to investigate the relation between the various glottographic waveforms and the vocal fold vibratory pattern.

The remainder of this chapter describes approaches used for laryngeal research, the anatomy and physiology of the larynx, and previous efforts to model the vibratory behavior of the vocal folds.

Approaches for Voice Excitation Research

Several approaches have been adopted for studying the phonatory mechanism. One approach monitors the mechanical activity of the vocal folds with direct or indirect observation techniques [Kitzing, 1986]. Some direct imaging techniques have been applied to investigate the vibratory patterns of the vocal folds [Moore, White, and von Leden, 1962]. Stroboscopic photography, used by Lecluse, Brocaar, and Verschuure [1975] and Teaney and Fourcin [1980], captured

the gross structure and movements of the vocal folds. However, details of the rapid vibratory motion of the vocal folds could not be observed with stroboscopic photography due to large intervals between successive strobe images. High-speed laryngeal photography, which overcame this disadvantage, captures vibratory behavior in detail [Moore, White, and von Leden, 1962; Hirano and Koike, 1974; Childers, Naik, Krishnamurthy, and Moore, 1983]. High-speed laryngeal film analysis is an important tool in laryngeal research, but there are several limitations for clinical practice. The technique is invasive and thus inconvenient for many patients; the post-processing of the film to extract information is time-consuming; and sounds produced by the subject are restricted by the procedure. Due to the relative inaccessibility of the larynx, it is difficult to observe directly the vocal fold vibration in vivo. Therefore, researchers resort to various indirect observation techniques such as PGG, ultrasonoglottography (UGG), EGG, and inverse filtering of the acoustic speech waveform (See Table 1-1). All indirect techniques record the variation of some physical quantity due to the vibratory motion of the vocal folds, and thus usually reflect different aspects of laryngeal behavior. The physical quantity can be the electrical impedance (EGG), the light intensity through the glottis (PGG), or the ultrasonic energy (UGG). Compared to direct observation techniques, indirect techniques are less invasive, less

impairing to continuous and fluent speech production, and easier to use. Despite these advantages, additional research is needed to model more accurately and interpret the signals from these indirect techniques.

Table 1-1 Widely used glottographies and their basic principles.

Glottography	measured physical quantity	related glottal signal
EGG	electric impedance	lateral contact area
PGG	light intensity	glottal area
UGG	ultrasonic energy	lateral contact area
Inverse-filtering of speech waves		glottal volume velocity

The second approach is to use an excised larynx mounted within an apparatus to supply airflow and to control the adjustments normally produced by the laryngeal muscles [Baer, 1975; van den Berg and Tan, 1959]. The setup has the advantage of being accessible for observations and measurements that are impossible with the intact larynx. There are, however, limitations to the use of excised larynges. Changes occurring in non-living tissues no doubt effect their mechanical properties somewhat. In addition, activity of the laryngeal muscles cannot be simulated adequately.

The third approach for the study of laryngeal behavior is quantitative modeling. Modeling efforts range from acoustic modeling of the vocal source to bio-mechanical modeling of the phonatory mechanism. The latter form of modeling explores the nature of self-oscillation of the human vocal folds. Acoustic modeling, often designed for voice synthesis, is usually aimed at parameterizing waveforms of the voice source. For linear prediction synthesis, it was found that the use of multiple pulses [Atal and Remde, 1982] or coded innovation sequences [Schroeder and Atal, 1985] to represent the LP residual yields improved speech synthesis over simple impulse excitation. Basically, both representations of the LP residual are adjusted so that the synthetic speech would be perceptually identical with the original speech, making the LP synthesis an excellent speech-copying method. However, its lack of physiological interpretation makes it difficult to apply the LP synthesis method to a specific voice quality. The voicing source used in formant synthesis has evolved from the simple sawtooth waveforms and filtered impulse trains to complex stylized glottal waveforms. Several stylized source excitation models have been proposed over the years [Rosenberg, 1971; Fant, 1979a; Ananthapadmanabha, 1984; Fant and Liljencrants, 1985]. These models use mathematically simple functions such as sinusoidal, exponential, polynomial waveforms, to approximate the waveshapes of true glottal area or glottal volume velocity functions. A well-designed

stylized glottal excitation model should allow direct control over the appropriate variables that are thought to be crucial for the quality of the synthesized voice.

On the other hand, the bio-mechanical models describe the laryngeal activity in terms of tissue properties and aerodynamic properties of the air stream. Although the general principle of the phonatory mechanism has been established, the formulation of a complete model of vocal fold vibration is complicated by several factors: an irregular, three-dimensional geometry, non-linear mechanical properties of the tissues (which are variable from place to place within the vocal folds), and various fluid mechanical phenomena in the air stream. Furthermore, the complex vibratory patterns change with the vocal register, pitch, intensity, health condition, and vocal tract configuration. How are the many varieties of phonation achieved? Irregularly shaped vocal fold boundaries and layers of tissue with vastly different mechanical properties, presumably altered by both extrinsic and intrinsic laryngeal muscles, seem to provide the necessary flexibility. Therefore, there is clearly a need for quantitative models exploring the detailed laryngeal behaviors governed by the spatially varying properties and the irregular three-dimensional shapes of the human vocal folds.

The Structure and Control of the Vocal Fold as a Vibrator

This section gives a brief introduction to laryngeal anatomy and phonatory physiology, describing what is known about the phonatory apparatus and its muscular control for speech production. This knowledge is important for modeling.

The operation of the speech production system as a whole is most readily divided into two functions, excitation and modulation. Excitation takes place mostly at the glottis but also at other points; modulation is achieved by the various organs of the vocal tract.

There are two ways for the glottis to convert an air stream into audible sound waves for speech. The first method involves using the air pressure to set the elastic vocal folds into vibration, producing a periodic sound wave. The second method involves placing a constriction at the glottis or supra-glottally, giving rise to turbulence, as in whispering and production of some voiceless aspirated plosives. The first method is termed voicing or phonation, and it is this vocal mode and its variation that we now consider. Since vocal sound and the phonatory disorders are generated in the larynx, this organ offers an appropriate place to begin a detailed description of the mechanisms of voice production.

Structure of the Larynx

The larynx is suspended from the hyoid bone and sits on top of the trachea. The larynx is a tube composed of

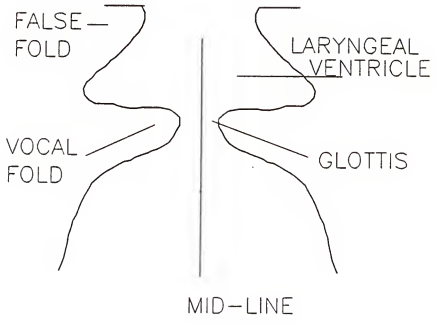


Figure 1-1 The frontal view of a human larynx [after Moore, 1971].

cartilages connected by ligaments and connecting membranes and covered by a mucous membrane. The enclosed area forms an hour glass space (Figure 1-1) with a vestibule above two sets of folds. One set of folds is the ventricular or "false" vocal folds. The other set is the "true" vocal folds which is used for voicing [Moore, 1971].

The ventricular folds form a second constriction, just above the true vocal folds. The vertical space between the two sets of folds is called the laryngeal ventricle and the horizontal space between the true vocal folds is called the glottis. Below the vocal folds the space widens again within the cartilaginous framework.

Laryngeal Cartilages

Three laryngeal cartilages form the basic frame within which the muscular control of phonation is exercised. These are the thyroid, the cricoid, and the paired arytenoid cartilages. Figure 1-2 is a schematic diagram of the relative position of these cartilages [Romanes, 1978].

The thyroid is a large, shielding cartilage that protects the front and sides of the larynx from injury. The muscles which make up the true and ventricular folds are attached to the front internal surface of the thyroid (Figure 1-3) [Laver, 1980].

The cricoid cartilage, so named because it is shaped like a signet ring, lies immediately below the thyroid. Although

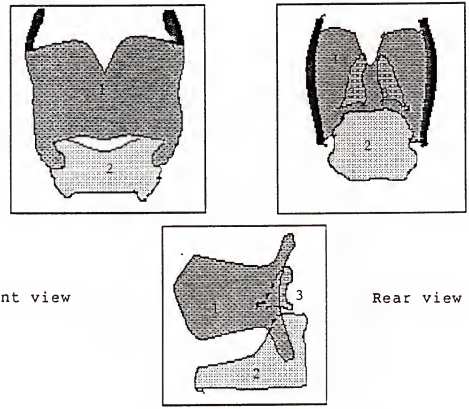


Figure 1-2 Schematic diagram of the principal laryngeal cartilages [after Romanes, 1978].
1. Thyroid cartilage 2. Cricoid cartilage
3. Arytenoid cartilage

the vocal folds are not attached to the cricoid cartilage, the cricoid cartilage articulates with three cartilages which do support the vocal folds.

The arytenoid cartilages, much smaller than the thyroid and the cricoid, are shaped like small pyramids and sit on the upper rim of the 'signet' of the cricoid at the back. The posterior ends of the true vocal folds are attached to the lower, forward angles of the arytenoid, called the vocal processes (Figure 1-3).

Muscles of the Larynx

The muscles of the larynx can be divided into an extrinsic and an intrinsic group. The extrinsic muscles are those which have their origins outside the larynx and attach onto various parts of the cartilaginous framework. They are mainly responsible for gross movements of the larynx as a whole and only affect the vocal folds indirectly. The intrinsic muscles, on the other hand, have both origins and points of insertion within the larynx. These muscles are directly responsible for modifying the configuration and tension of the vocal folds. The intrinsic muscles achieve these changes by movement of the cricoid about the cricothyroid joint, by the rocking and gliding motion of the arytenoid cartilages, and by isometric contraction of muscles with the vocal folds themselves. Since muscles function actively in only one way (they shorten by contraction), there are opposing groups of

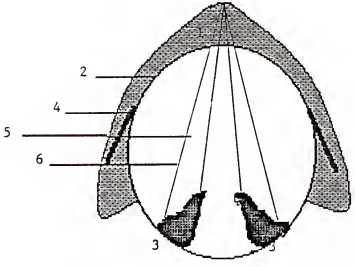


Figure 1-3 Schematic diagram of the laryngeal muscles connecting the arytenoid cartilage to the thyroid cartilage [after Laver, 1980].

1. Thyroid cartilage
2. External edge of cricoid
3. Arytenoid cartilage
4. Cricothyroid muscle
5. Ventricular fold
6. True vocal fold

muscles that draw their associated structure in one direction or the other. The extrinsic muscles can be divided further by the way they position the larynx into laryngeal elevators and laryngeal depressors. Similarly, the intrinsic muscles can be divided according to their adjusting functions into abductor, adductor, tensor, and relaxer [Fink and Demarest, 1978].

Longitudinal tension of the true and ventricular vocal folds can be achieved by two different actions. The first action is that of retraction and slight vertical rotation of the cricoid by means of the cricothyroid, which puts longitudinal tension on the vocal folds by stretching them. The second action is contraction of the muscles that make up the vocal folds themselves, the thyroarytenoid (vocalis muscles).

Those muscles that control the positions of the arytenoid cartilage relative to the cricoid, have the function of opening and closing the glottis (abductor and adductor, respectively).

The Vocal Folds

Physiologists have isolated at least three distinct elements of the vocal folds that have functional roles in phonation: (1) the vocal ligament, (2) vocalis muscle and (3) mucous membrane (mucosa). A sketch of these elements is shown in Figure 1-4 [Titze and Strong, 1975].

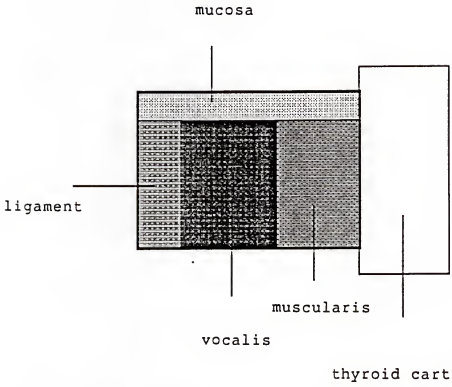


Figure 1-4 Simplified sketch representing the internal tissue structure of the left vocal fold in frontal cross-section [after Titze and Strong, 1975].

The vocalis is the main body of the vocal folds. It consists primarily of longitudinal muscular fibers whose tensions are variable. The vocal ligament supports tension applied externally when the vocalis muscle is relaxed. It is attached close to the vocalis muscle on the side of the glottis. The mucous membrane covers the vocalis muscle and the vocal ligament.

Models of Vocal Fold Vibration

The purpose of this section is to discuss the development of models of the phonatory mechanism and to show how their limitations led to the present research.

The Aerodynamic-Myoelastic Theory

A widely accepted theory of vocal fold vibration is the aerodynamic-myoelastic theory. Van den Berg [1968] postulates in the aerodynamic-myoelastic theory that the function of the larynx is based on the interplay of three factors. First, aerodynamic properties of the air actuates the larynx. Second, the larynx is adjusted by the neural activation of the various muscles and the myoelastic properties of the laryngeal components. Third, the subglottal system and the larynx, the left and right vocal folds, and the larynx and the supraglottal system are coupled aerodynamically.

The different aerodynamic coupling factors is important in considering the fine details of vocal fold vibration. The

detailed mode of vibration of the vocal folds will depend partly on the degree of effort exerted sub-glottally by the pulmonary system so that characteristics of vibration will necessarily co-vary with subglottal pressure. Also, coupling factors between the left and right vocal folds become important perceptually when organic asymmetries of the larynx are considered, as in the cases of nodules and polyps on either vocal fold. In addition, the articulatory state of the supraglottal vocal tract will influence the airflow out of the glottis. This source-tract interaction will even influence the fine detail of the vibratory pattern of the vocal folds to some small degree.

According to this theory, the glottal airstream and the vocal folds form a mechanical system that may demonstrate instability for oscillation under specific flow conditions. The aerodynamic-myoelastic theory, however, does not offer a mathematical foundation to explain the interaction between the glottal airstream and the elastic laryngeal tissues and conditions under which flow induced vibrations can be sustained. Ishizaka and Matsudaria [1972] provided the analysis of the aerodynamic forces and the simplest mechanical structure, the two-mass representation of the vocal folds, which is capable of responding to these forces.

Two-Mass Model of the Vocal Fold Vibration

In the two-mass model, the vocal folds are divided in depth into an upper and a lower mass due to the anatomical and functional division between the mucosa and the vocalis. Each part consists of a simple mechanical oscillator having mass, spring, and damping (m, s, r); see Figure 1-5 [Ishizaka and Flanagan, 1972]. The springs represent the elastic properties of the folds, while the damping represents dissipative forces such as viscosity and friction. There is also an interaction between the two masses, represented by a coupling stiffness S_c. The coupling stiffness represents the fact that as one of the masses is displaced relative to the other, there is a force tending to restore the masses to their equilibrium position relative to one another.

Computer simulation results. In addition to quantifying the interaction between the elastic vocal folds and the aerodynamic forces, a computer simulation based on this model has also successfully captured many important features of the behavior of human vocal folds. It showed that (1) the pre-phonatory area, or rest area of the vocal folds is a critical factor in establishing self-oscillation, (2) regions of stability and instability of the entire system could be bracketed on the basis of average air flow, glottal opening, and spring constants, (3) acoustic interaction between the vocal tract configuration and the glottal volume velocity is strong, and (4) the acoustic parameters of the voice, such as

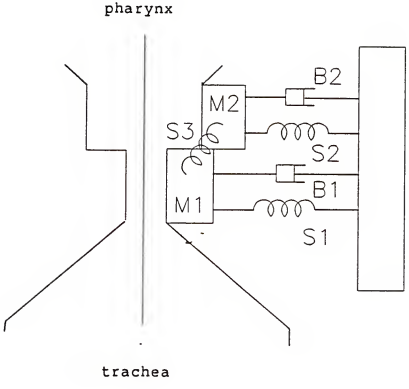


Figure 1-5 The two-mass model of the vocal folds [after Ishizaka and Flanagan, 1972].

fundamental frequency, intensity, and register, are influenced by physiological adjustments such as subglottal pressure, pre-phonatory glottal opening, and tension of the vocal folds.

Multiple-Mass Model

Although the two-mass model served as a milestone in quantifying vocal fold vibration, it only modeled the vocal folds as a minimal mechanical structure capable of responding to aerodynamic forces and sustaining oscillation. It was not capable of exhibiting various longitudinal vibratory modes observed in human phonation. Among these are vibrations in which glottal closure occurs only along certain portions of the glottis, at different times for different positions (longitudinal phase differences), or in which the folds can exhibit complicated motion patterns due to horizontal inhomogeneities such as vocal nodules and polyps. Titze [1973], in an attempt to enlarge the horizontal degree of freedom, proposed a 16-mass model composed of two rows of eight masses each (Figure 1-6). The top-row of masses represents primarily the mucosa and the bottom row represents primarily the vocal ligament and the vocalis muscle. The forces T_m and T_v represent the longitudinal tensions as determined by the balance of forces between the cricothyroid and thyroarytenoid muscles. Specifically, the spring constants for the upper and lower rows increase nonlinearly with elongation of the vocal folds. The 16-mass model has

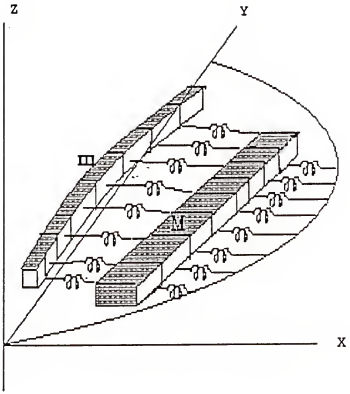


Figure 1-6 A 16-mass model of the vocal folds [after Titze, 1973].

conditions for sustained oscillation analogous to those in the two-mass model.

Continuum Model

One could, of course, extend the above model by increasing the number of masses and degrees of freedom. Titze and Strong [1975] went a step further than this when they represented the vocal folds not as a coupled set of discrete masses but as a continuous deformable medium (Figure 1-7). In addition, the incompressibility of the vocal folds dictates a coupling between horizontal and vertical motion. An important consequence of the incompressibility of the vocal folds is that the vibratory mode most easily excited appears to involve vertical phase differences, since this mode tends to preserve the volume of the vocal folds. Their study also showed that the layered structure of the vocal folds is ideally adapted to support vocal fold vibration. The longitudinal fibrous structure, for example, is much more loose in the vertical direction than in the longitudinal direction. This allows vertical phase differences to occur.

The continuum model was highly informative about the relation between the vocal fold structure and the vocal fold vibratory mode. However, the shape of the vocal folds in this model was restricted to a rectangular form. The tissue properties were uniform in the plane normal to the longitudinal direction for ease of manipulation [Titze, 1976].

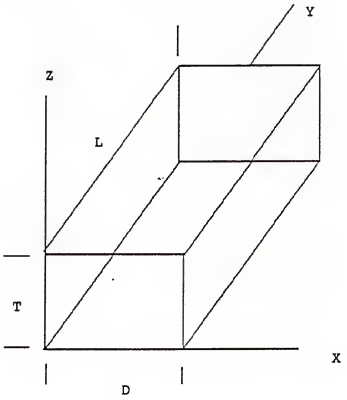


Figure 1-7 A continuum model of the vocal folds [after Titze and Strong, 1975].

In addition, the model lacks a complete representation of the interaction between the aerodynamic air flow and the elastic vocal fold tissue because the normal modes of vocal fold vibration were derived based on an eigen-value analysis of the fold tissue.

Glottographic signals are widely used for assessment of laryngeal function and for speech analysis. However, none of the models discussed above was intended to generate realistic glottographic waveforms. To more accurately model and interpret glottographic waveforms, a proper representation of irregularly shaped vocal fold boundaries and layers of tissue with different mechanical properties should be included into models. Along this line, Childers, Hicks, Moore, and Alsaka [1986], Titze [1984 and 1988c], and Titze and Talkins [1979b] proposed their models for glottographic signals. In the next chapter, we will demonstrate our approach to this problem.

CHAPTER 2

A KINEMATIC MODEL OF VOCAL FOLD VIBRATION

The Basic Objectives of the Modeling

The first aim of the model is to characterize the vocal fold vibratory characteristics in terms of a 3-dimensional glottal configuration with spatially varying tissue properties. The second aim is to model several glottographic waveforms. Because of the wide use of indirect techniques in monitoring the laryngeal vibratory activities and in assessing the vocal source function, a general mathematical formulation that predicts glottographic waveforms from specific vocal fold configurations is desired.

The glottographic waveforms obtained with indirect techniques have been related to such vocal fold events as glottal area, vocal fold contact area, and glottal air flow. Therefore, interpretation of glottographic signals requires two stages of research. For example, the first stage might determine the extent the EGG represents vocal fold contact area and the PGG represents glottal area. In the second stage, the relations between vocal fold contact area and glottal area, and vibratory patterns of the vocal folds might be established. The research reported here deals with the

modeling of several vocal fold events and the relation of these events to the modeled glottographic waveforms.

Development of the Vibratory Model

Choice of Model Type and Model Parameters

This subsection describes how the structure and the parameter set of the proposed model were chosen.

The proposed model is, as was Titze's vibratory model [1984], basically aimed at describing the kinematics of vocal fold vibration. Compared to Titze's vibratory model, the proposed model is more flexible in describing a 3-dimensional configuration of human vocal folds; it incorporates more vibratory features, and allows simulations for an irregular vibratory pattern and asymmetric fold properties.

Some aspects of vocal fold vibration are interpreted as travelling wave phenomena. The outer tissue layer of the vocal folds is assumed to propagate a surface wave along the length of the glottis and in the direction of air flow [Hirano, 1975]. To describe such wave-like motion, we assume that each point on the surface of the vocal folds has the same time-dependent displacement function, as in other types of wave propagation. Furthermore, the maximum excursion of the vocal folds and equilibrium positions may vary according to the glottal configuration.

Since the model requires the displacement function of the vocal fold tissue to be an input variable, the model does not

explore the self-oscillatory nature of the vocal folds. Nevertheless, its ability to describe the vibration of 3-dimensional vocal folds makes the model a powerful tool to study glottographic signals.

Since each glottographic waveform reflects an average measurement of the vocal fold motion, it is natural to formulate the glottographic waveforms in terms of vocal fold kinematics. The choice of model parameters is based on their importance in developing and adjusting the self-oscillation of the vocal folds and their appropriateness in describing the kinematics of the vocal folds.

In essence, model building is a systematic coordination of theoretical and empirical elements of knowledge into a joint construct. Empirical observations of vocal fold vibration enter at this stage of model construction by suggesting specific information about the glottal configuration and vocal fold kinematics that should be contained in the parameter set. Ultra-high speed photography of the vibrating folds [Moore and Von Leden, 1958; Hirano, 1983] and the results from previous vocal fold models [Ishizaka and Flanagan, 1972; Titze and Talkin, 1979a] have demonstrated that five factors are important in determining vibratory patterns:

(1) Proper abduction and adduction of the vocal folds. It is well known that proper abduction must be achieved before vibration. Different levels of abduction may also result

in different vibratory modes [Ishizaka and Flanagan, 1972];

(2) The vertical pre-phonatory shape of the glottis. Titze [1988a] has shown that the vertical shape may be of significance in register control in human phonation and in determining waveshape of the glottographic waveforms;

(3) Mucosal wave, also known as vertical phasing. A lag between the movements of the upper and lower portions of the vocal folds has been observed during phonation, except for falsetto. This phenomenon is more obvious during low pitch phonation. Photography of the vibration of the vocal folds in the frontal plane confirmed the wave-like nature of the vibration [Moore and Van Leden, 1958]. This lag reflects the degree of coupling along the depth of the vocal folds;

(4) A lag between the movements of the anterior and the posterior portions of the vocal folds, known as longitudinal phasing. During the opening phase, the vocal folds first separate along their posterior halves and the separation progresses toward the anterior halves [Hirano et al., 1983];

(5) The profiles of maximum excursion of the vibration along the length and the depth of the glottis. The maximum displacement attained from its equilibrium position is called the amplitude of oscillation or maximum excursion.

These distributions are related to the non-uniform tissue properties in each specific dimension.

Assumptions for Modeling

Several assumptions have been made, without the loss of any key feature, to simplify the mathematical expressions in the proposed model of vocal fold vibration.

1) The vibratory movement of vocal folds is confined to the lateral direction. Although motion of the fold tissue does occur in other directions, especially after collision between the two folds, the projected area of the glottis is primarily determined by the vocal fold lateral motion. As for changes in lateral contact area due to vertical movement when the folds collide, an algorithm is proposed to take this into account.

2) At the first stage of simulation, each point on the medial fold surface vibrates in a sinusoidal fashion, i.e., as a harmonic oscillator [Titze, 1984]. This is a first-order approximation to the true displacement function of vocal folds. Despite its simplicity, the model based on this approximation can simulate most features of the glottographic waveforms. On the other hand, the vibratory pattern of the vocal folds for abnormal voices may be quite irregular and other surface vibratory functions need to be examined.

3) For all travelling wave phenomena in the proposed model, uniform travelling speed is assumed. Thus, the time delay in movement of different points on the fold surface is proportional to their location in the direction of wave propagation.

Description of Vocal Fold Kinematics

For the sake of the reader, a pre-phonatory vocal fold configuration is shown in a cartesian coordinate system in Figure 2-1. The length of the glottis is parallel to the Y axis, the width to the X axis, and the air stream flows in the direction of the positive Z axis.

Conceptually, the displacement function of the glottis is decomposed into a static component and a dynamic component. The static component represents the 3-dimensional pre-phonatory configuration of the glottis and the dynamic component represents the time-varying part during oscillation.

For computer simulation, the medial surfaces of the vocal folds are divided into several small elements and each element is treated as an oscillator. To describe the motions of a large number of oscillators, one needs to specify the equilibrium position (Figure 2-1) and maximum excursion of vibration (Figure 2-2) for all oscillators on the medial surfaces along with the spatial phase relation between these oscillators. The pre-phonatory configuration of the glottis,

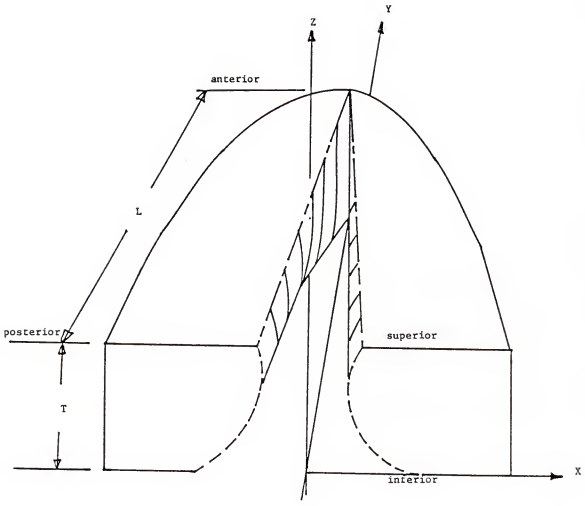


Figure 2-1 A pre-phonatory configuration of the vocal folds [after Titze, 1984].

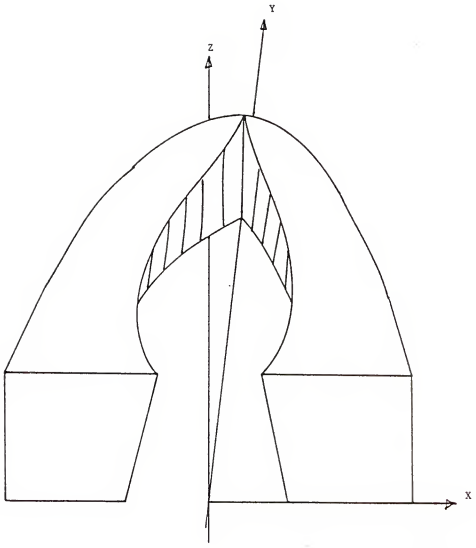


Figure 2-2 A maximum excursion profile of the vocal fold. Along the length of the glottis, amplitude of vibration is zero at the anterior end, and attains maximum in the middle. Along the depth of the glottis, the inferior end has a larger amplitude than the superior end does.

depicted in 3-dimensional form in Figure 2-1, establishes the equilibrium positions for each oscillator.

The pre-phonatory configuration of the medial surfaces and the maximum excursion profiles need be specified on the frontal plane (X-Z plane) as well as on the horizontal plane (X-Y plane). After the pre-phonatory configuration, the maximum excursion profile, and the spatial phase relation have been specified, the displacement functions for elements on the medial surfaces of the vocal folds will be:

$$\text{Disp}(y,z) = \text{pre-phonatory configuration}(y,z) + \text{amplitude}(y,z) * \xi(t-(z/c_z)-(y/c_y)) \quad (2\text{-}1)$$

where ξ is the dynamic component of the displacement function, t the time, c_z the propagating velocity along the depth, c_y the velocity along the length. In Titze's formulation [1984], glottal displacement was also decomposed into a static component and a dynamic component. In his model, however, the vertical shape of the glottis must be trapezoidal; the maximum excursion profile is fixed; and the lag between movements of the anterior and the posterior portions of the vocal folds was not taken into account.

Substituting the relation, c = distance / duration, into equation (2-1), we have:

$$\text{Disp}(y,z) = \text{pre-phonatory configuration}(y,z) + [\text{amplitude}(y,z) * \xi(t-(z*T_z/\text{thickness}) - (y*T_y/\text{length}))] \quad (2\text{-}2)$$

where T_z: lag between the superior and the inferior margins.

T_y: lag between the anterior and the posterior margins

y: index along the length of the glottis

z: index along the vertical axis (Z+ is the direction of air flow.)

amplitude(y,z): the amplitude assigned to the element at (y,z) on the medial surfaces

pre-phonatory configuration(y,z): the pre-phonatory position at the location (y,z) on medial surface.

Nominal Glottal Configurations and Physical Constraints

Feasible glottal configurations are limited by anatomy and physiology of the larynx and its muscular control. The following is a brief discussion on the constraints for glottal configuration and model parameters.

Pre-phonatory configuration of the glottis. The shape of the glottis depends on the structure surrounding the vocal folds and the muscular adjustments of the structure. The anterior end of the vocal folds is fixed to the thyroid cartilage while the posterior end is attached to the movable arytenoid cartilages. Thus, the gross horizontal configuration of the glottis forms an isosceles triangle with the anterior end of the vocal folds at its vertex. The vertical shape of the glottis is usually trapezoidal and can vary in shape from a funnel (divergent) to a cone (convergent) [Titze, 1984].

There are variations from the nominal configuration, however, where outlines can be curves rather than straight lines. In addition, the boundary can be of irregular shape due to the presence of a tumor or polyp.

Amplitude profile of vocal fold vibration. The maximum excursion profile reflects the degree of compliance of the fold tissue along each specific dimension. Along its length, the vocal folds appear to be most pliable near the mid-point of the membranous portion. The posterior portion of the vocal folds appears to be slightly more pliant than the anterior portion. At the anterior and the posterior ends, the tissues appear to be firm. The largest amplitude of lateral excursion, however, occurs not at the midpoint but at a place posterior to the midpoint [Hirano, 1983]. This can be accounted for, at least in part, by the fact that the anterior end of the vocal folds is a definite fixed point, while the posterior end is movable.

An additional vibratory parameter, DCAC, determines the ratio of the average abduction to the average amplitude of the oscillators [Titze, 1984]. The DCAC varies according to voice modes and individual discrepancy. It has, for example, a larger value in falsetto register, in which the vocal folds are more abducted and the amplitude of vibration is smaller than in modal register [Laver, 1980].

Wave propagation speed and phase difference. The lags in movement along the length and the depth of the vocal folds

can be expressed in the model as a fraction of a pitch period of voicing. The velocity of propagation in the direction of air flow varies from 1.5 to 2 meters/sec [Baer, 1975]. At a fundamental frequency of 100 Hz and for a fold thickness of 2 mm, the lags can be translated into 48° and 29°, respectively. Timcke, von Leden, and Moore [1958] described a vertical phase difference of 43 degrees. The computer simulation of the flow-separation theory by Ishizaka and Flanagan [1972] resulted in a typical phase difference of 55 degrees.

On Setting of Model Parameters

Once the structure of the vibratory model is selected, a new question arises: How do we choose proper model parameters for various phonating conditions? The choice of model parameter values is presently based on empirical data, with each parameter setting being made separately. In fact, parameters such as vertical phase difference, vertical shape of the glottis, and fundamental frequency of vibration are mutually correlated and should not be assigned arbitrarily. Knowledge of such correlations would help us exclude physically impossible combinations of parameters. To develop a more systematic method for parameter settings, we propose a modified two-mass model (Figure 2-3) to explore combinations and ranges of the parameter values. The approach is to vary systematically control input, e.g., lung pressure, fold

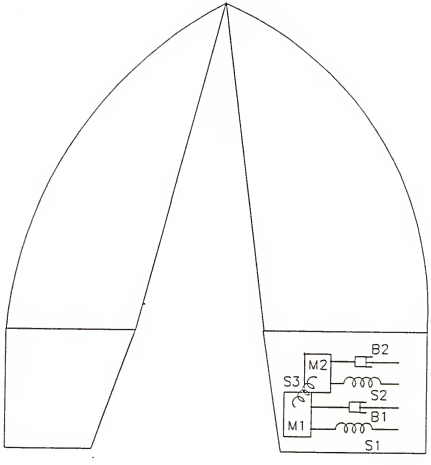


Figure 2-3 The modified two-mass model.

tension, or abduction of the glottis, of the modified two-mass model and determine the corresponding changes in kinematics of the vocal folds.

In the modified two-mass model, the displacements of the upper and the lower masses in the original two-mass model [Ishizaka and Flanagan, 1972] are assumed to be the displacements of the upper and the lower edge of the vocal folds. The displacements of elements between these two edges are derived by interpolation. The horizontal configuration of the glottis is chosen as an isosceles triangle with the anterior end of the vocal folds being its vertex and the length of the baseline varying according to the degree of abduction.

The input parameters of the modified two-mass model are the pre-phonatory vertical configuration, lung pressure, and fold tensions. The pre-phonatory configuration is specified by the abduction of the upper and the lower edges (x_{2_0} and x_{1_0}, respectively) near the posterior end. If x_{1_0} is greater than x_{2_0}, then the vertical configuration is convergent; otherwise it is divergent.

The Two-Mass Model

As described in Chapter 1, the motions of the masses in the two-mass model [Ishizaka and Flanagan, 1972] are governed by the aerodynamic forces that activate the larynx and the myoelastic forces of the springs and the dampers. Here we

take a closer look at the functions of various parameters in the two-mass model.

The basic equations of motion are [Ishizaka and Flanagan, 1972]:

$$m_1 d^2x_1/dt^2 + r_1 dx_1/dt + k_1 x_1 + k_c(x_1 - x_2) + F_1 = 0 \quad (2\text{-}3)$$

$$m_2 d^2x_2/dt^2 + r_2 dx_2/dt + k_2 x_2 + k_c(x_2 - x_1) + F_2 = 0 \quad (2\text{-}4)$$

where x_1 and x_2 are the lateral displacements of the lower and upper mass respectively; m, r, and k represent mass, viscous loss, and spring constant, respectively; and F_i is the aerodynamic forces exerted on each mass. In the simulations, the springs are given a nonlinear characteristic to conform to the stiffness as measured on fresh, excised human vocal folds [Ishizaka and Flanagan, 1972]. During closure of the glottis, there is a contact force which results in additional deformation. The spring restoration forces are represented as [Ishizaka and Flanagan, 1972]:

$$f_{si} = k_i x_i (1 + \eta_{ki} x_i) \qquad i = 1, 2 \quad (2\text{-}5)$$

$$f_{hi} = h_i (x_i + x_{0i})(1 + \eta_{hi}(x_i + x_{0i})) \quad (2\text{-}6)$$

$i = 1, 2$ for $x_i + x_{0i} < 0$

where k and η_k are linear and nonlinear stiffness of the springs respectively; h and η_h are additional linear and nonlinear stiffness of the springs, respectively, during the closed phase. The viscous resistances are [Ishizaka and Flanagan, 1972]:

$$r_i = 2\zeta_i (m_i k_i)^{1/2} \quad i = 1, 2 \quad (2\text{-}7)$$

where ζ is the damping ratio.

The physiological constants shown above are set to values suggested by Ishizaka and Flanagan [1972]; namely:

$m_1 = 0.125$ g	lower mass
$m_2 = 0.025$ g	upper mass
$d_1 = 0.25$ cm	thickness of m_1
$d_2 = 0.05$ cm	thickness of m_2
$l = 1.5$ cm	length of vocal folds
$k_1 = 80000$ dyne/cm	linear stiffness of spring 1
$k_2 = 8000.$ dyne/cm	linear stiffness of spring 2
$\eta_{k1} = \eta_{k2} = 100.$ dyne/cm	nonlinear stiffness of s_1 and s_2
$h_1 = 3 * k_1$	linear stiffness of s_1 for contact spring
$h_2 = 3 * k_2$	linear stiffness of s_2 for contact spring
$\eta_{h1} = \eta_{h2} = 500.$ dyne/cm	nonlinear stiffness of s_1 and s_2 for contact spring
$k_c = 25000$dyne/cm	stiffness of coupled spring
$x_1 = 0.1$	damping ratio for open phase
$x_2 = 0.6$	
$x_1 = 1.1$	damping ratio for closed phase.
$x_2 = 1.6$	

The lower mass, the vocalis-ligament combination, is five times as massive and five times as thick as the upper mass, the mucous membrane.

For simulation of normal voices, these parameters are held constant at the above values, except for control of fundamental frequency. Ishizaka and Flanagan [1972] proposed a "tension parameter," Q, to control fundamental frequency. For the range of 120 to 220 Hz, the fundamental frequency varies almost linearly with the tension parameter Q. For other vocal registers such as falsetto and vocal fry, however, the parameter setting above is no longer valid. Similar problems occur when attempts are made to simulate the vibratory pattern of abnormal vocal folds. Therefore, the

parameters in the two-mass model should be changed according to laryngeal adjustments for other voicing modes.

Inter-relation Between Model Parameters

In this section, variations of and the inter-relation among kinematic parameters of the proposed model are investigated based on the modified two-mass model, as the physiological parameters, (e.g., pre-phonatory configuration of the glottis, fold tension, and lung pressure), are changed. The kinematic parameters to be investigated are: fundamental frequency, vertical phase difference, vertical equilibrium shape, vertical amplitude profile, and maximum excursion of vibration. The use of the two-mass model also limits the kinematic parameters we can study with this approach. The variations of vocal fold dynamics along the length of the glottis cannot be simulated with this approach because the two-mass model can only account for vertical inhomogeneity.

Varying the fold tension. The two-mass model [Ishizaka and Flanagan, 1972] is adjusted by scaling up all the spring constants and scaling down the masses and thicknesses by the factor Q. Figure 2-4 illustrates the behavior of the two masses at three different tension levels (Q = 0.8, 1.0, and 1.5). The results show that an increase in the tension factor Q will increase the fundamental frequency, while reducing the phase difference and the amplitude of oscillation. The masses vibrate in smaller amplitude as the stiffness of the spring

becomes greater. The decrease in vertical phase difference as a result of increased tension can be expected because the two-mass model behaves more like an one-mass model when the coupling spring becomes more rigid. An increase in Q also makes the upper mass oscillate in a greater amplitude than that of the lower mass. This trend leads to a more left-skewed glottal area waveform.

Varying the lung pressure. Figure 2-5 shows the behavior of the two masses at three different pressure levels [Ishizaka and Flanagan, 1972] (PS = 5000, 9000, and 15000 dyne/cm^2). Note that both the fundamental frequency and the amplitude of vibration are increased as subglottal pressure becomes greater. The increase in fundamental frequency with subglottal pressure might be ascribed to the nonlinearity of the deflection of the muscles and ligaments at large amplitudes of vibration [Ishizaka and Flanagan, 1972]. The large displacement amplitude, due to an increase in aerodynamic driving force, increases the effective stiffness of the springs, and consequently increases fundamental frequency. The slope of fundamental frequency as a function of subglottal pressure is about 2.5 Hz per 1000 dyne/cm^2 for hk = 100.

Varying the pre-phonatory configuration of the glottis. In human phonation the degree of abduction is controlled by those laryngeal muscles that position the arytenoid cartilage relative to the cricoid [Laver, 1980]. Three different

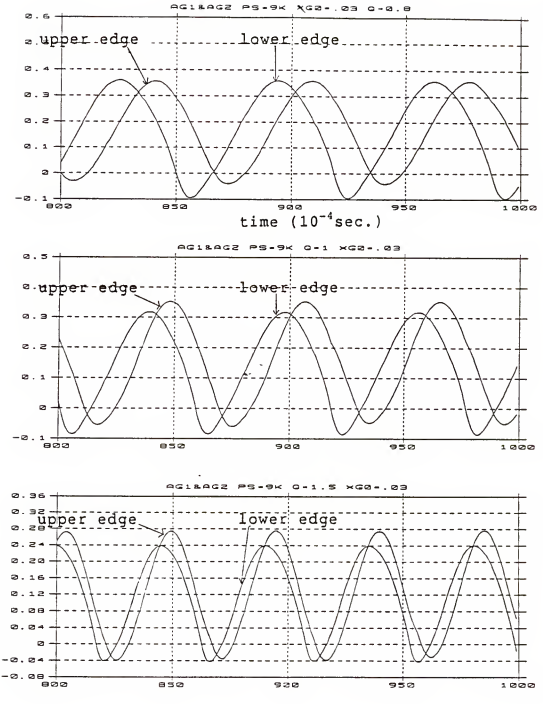


Figure 2-4 The lateral displacement of the upper and the lower edges of the vocal folds at three different tension levels (Q = 0.8, 1.0, and 1.5.)

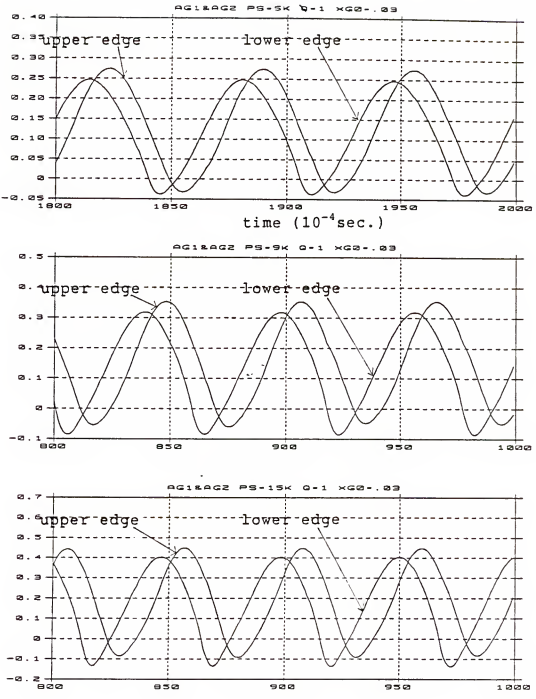


Figure 2-5 The lateral displacement of the upper and the lower edges of the vocal folds at three different subglottal pressure levels (P_s = 5000, 9000, 15000 dyne/cm^2)

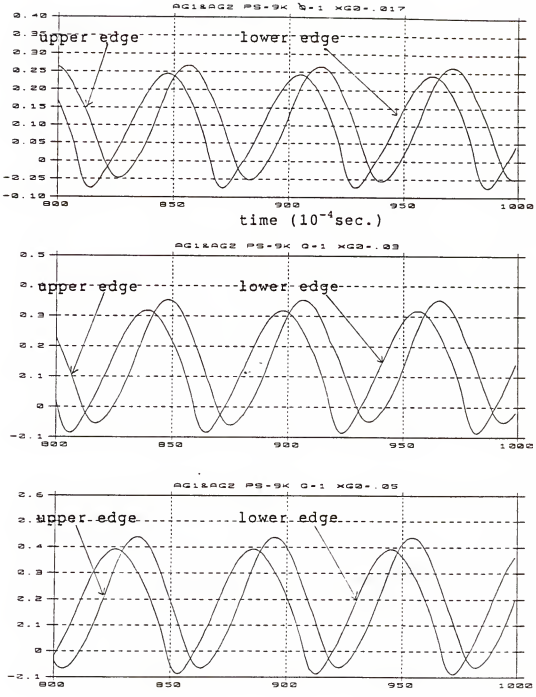


Figure 2-6 The lateral displacement of the upper and the lower edges of the vocal folds for three different degrees of abduction (0.017, 0.03, and 0.05 cm).

degrees of abduction (0.017, 0.030, and 0.050 cm) are simulated and the results are shown in Figure 2-6 [Ishizaka and Flanagan, 1972]. Like a perfect spring, of which the amplitude of oscillation is proportional to the initial distance from the rest position, the masses vibrate in greater amplitude when the degree of abduction is larger. Unlike a perfect spring, if the abduction exceeds some limit, the masses cease to vibrate, because pressure is not built up at the glottis and set the vocal folds into vibration.

In addition to the degree of abduction, the shape of the glottis prior to phonation also affects the behavior of the vocal fold model. Shown in Figure 2-7 is the behavior of the two masses for three pre-phonatory vertical shapes of the glottis, i.e., a divergent, a parallel, and a convergent shape. For a divergent configuration, the fundamental frequency is slightly higher; the amplitude of vibration of upper mass is greater; the displacement function of the lower mass is more asymmetric; the duration needed to build up vibration is shorter; and the vertical phase difference is less than that of a convergent configuration. It is interesting that changing the shape of the glottis alone could result in a change in the fundamental frequency. This suggests that the shape of the glottis might play a role in control of the vocal register.

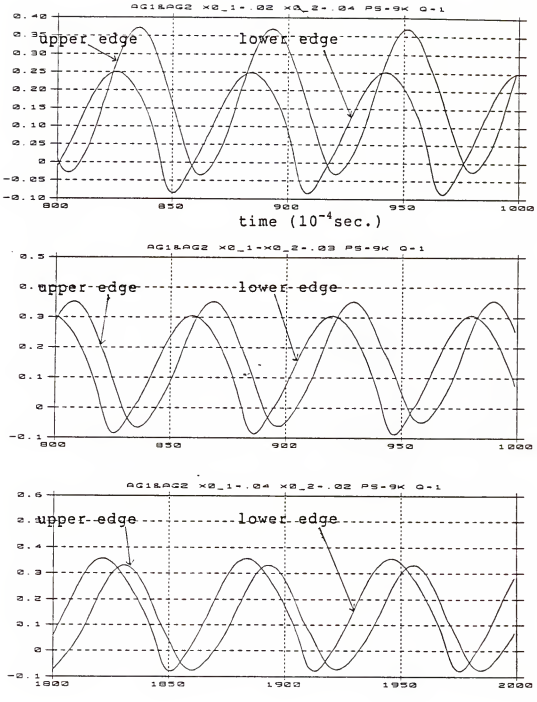


Figure 2-7 The lateral displacement of the upper and the lower edges of the vocal folds for three pre-phonatory vertical shapes of the glottis. The upper graph is for a divergent shape. The middle graph is for a parallel shape. The lower graph is for a convergent shape.

A Realistic Displacement Function of the Glottis

At first, we assumed that the displacement function of the elements on the medial surfaces of the vocal folds is a sinusoidal type. This assumption, however, is not valid for many vibrations of the vocal folds we observed in laryngeal films, especially for abnormal voices. In this section we consider why we need a more realistic displacement function of the vocal folds and present a solution to this need.

Simulation with the two-mass model may give us hints about a realistic displacement function for the glottis. In Figure 2-8 is the deviation of the simulated displacement function from a sinusoidal function. The waveform of the displacement of the lower mass skews slightly to the right and falls steeply at the closure. Both trends may contribute to the characteristic steep trailing-slope of a typical EGG waveform. On the other hand, the waveform of the displacement of the upper mass is symmetric and gradual at closure.

For some extreme cases, the displacement can deviate substantially from a sinusoid. Shown in Figure 2-9 are displacement functions for different laryngeal settings of the two-mass model. Note that waveshape symmetry and slope at the opening and the closure can vary from one laryngeal setting of the glottis to another. Hirano, Matsuo, Kakita, Kawasaki, and Kurita [1983] photographed, with indirect laryngoscopy, the vibration of the vocal folds of abnormal subject and registered displacements of 13 points along the length of the

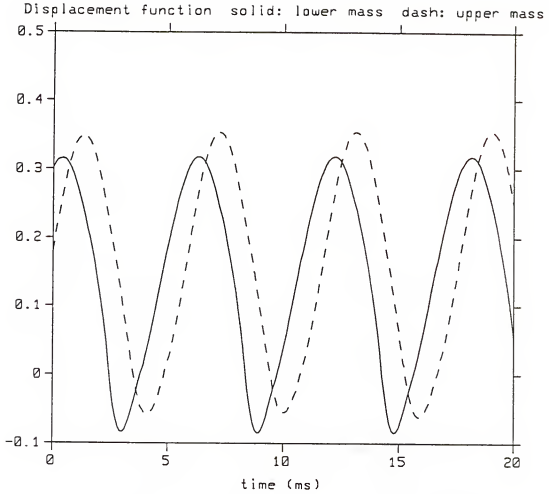


Figure 2-8 The lateral displacement of the lower (solid line) and the upper mass (the dashed line) in the two-mass model for a typical condition (subglottal pressure: 9000 dyne/cm^2, tension factor: 1, and abduction: 0.03 cm).

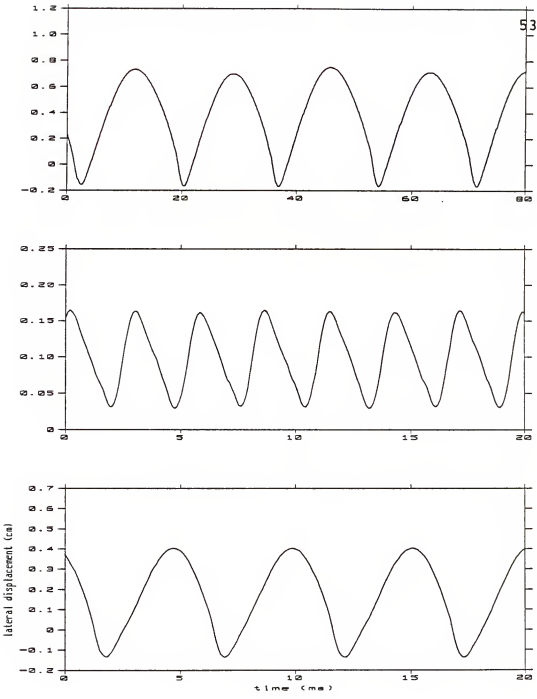


Figure 2-9 The displacement functions for different laryngeal settings in the two-mass model. The figures show that waveshape symmetry and slope at opening and closure can vary from one laryngeal setting to another. For the upper graph, Q = 0.3 and K_2 = 2000. For the middle graph, Q = 2.0 and K_2 = 80000. For the lower graph, Q = 1.0 and K_2 = 8000.

glottis. Tanabe [1975] showed the vibratory patterns of 5 points along the length of the vocal folds for 12 different phonations. Their measurements showed variations in symmetry and slope in the displacement of the vocal folds.

The observations above motivated us to develop a mathematical approximation which allows control over the symmetry and the slope of a sinusoid-like waveform. Aside from extreme cases, e.g., where the displacement function has two humps, it is possible to approximate the displacement function with a family of simple curves. We propose a mathematical expression to represent the displacement function of the glottis. Two parameters are used in this approximation: the speed quotient, SQ, which controls the symmetry of the waveform and the slope quotient, PQ, which governs the rising and falling slopes. The equation governing the displacement function is:

$$\begin{aligned} disp(t) &= 2*(SIN(t*p/2/\gamma)PQ+\log(SQ)-1/2) \quad 0 < t < \gamma \\ &= 2*(SIN((t-\gamma)/(T-\gamma)*p/2+p/2))PQ-1/2) \quad \gamma < t < T \end{aligned}$$

where $\gamma = SQ*T/(1+SQ)$

T : Pitch period

t : Time (2-8)

Basically the normalized displacement function is a distorted sine function. The way it is distorted is determined by the exponent PQ. Since $SIN(\theta)$ falls in the range [0,1] for $0 < \theta < p$, choosing PQ less than 1 will enlarge the amplitude at the closure, resulting in a gradual

slope near the closure. On the other hand, a function with PQ greater than 1 will compress the amplitude at closure and make a steep slope at closure. Some additional adjustments make the values of the displacement function fall in the range [-1, 1].

The SQ determines the symmetry of the waveform and is defined as:

$$SQ = \frac{\text{duration of positive-going segment}}{\text{duration of negative-going segment}} \qquad (2\text{-}9)$$

According to the definition, an SQ greater than 1 will make a right-skewed waveshape, while an SQ less than 1 will make a left-skewed waveshape.

Simulations using the two-mass model revealed that asymmetry in the waveform of a displacement function often results in differences in the slopes of rising and falling segments. Thus, a logarithmic term, log(SQ), is added to the exponent of the rising segment to model this phenomenon. The logarithmic term gives the rising segment a greater slope quotient if the waveform skews to the right. The displacement function is plotted in Figure 2-10 for various settings of the parameter SQ and PQ.

Summary

The purpose of this chapter was to develop a unique kinematic model that is capable of describing the complex motion of the vocal folds. The proposed kinematic model of

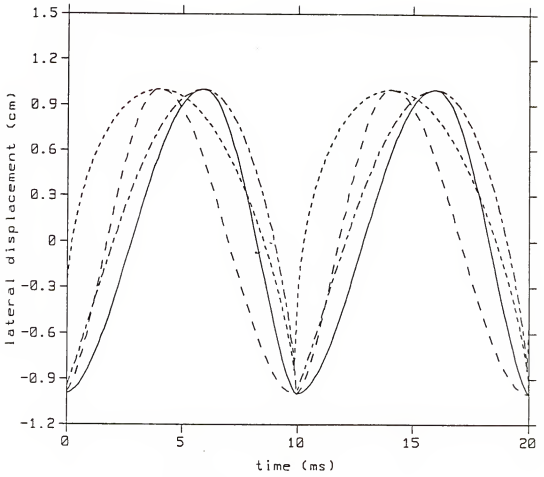


Figure 2-10 The displacement functions generated from the equation 2-8 for various settings of the parameter SQ and PQ.

	SQ	PQ
solid line:	1.5	1.5
dashed line:	0.7	2.0
dotted line:	0.7	0.7
dot-dash line:	1.5	0.7

vocal fold vibration has incorporated several concepts of vibratory behavior of the vocal folds. These are: 1) pre-phonatory configuration of the glottis, 2) maximum excursion profiles along the length and depth of the glottis, and 3) time lags in movement along the length and depth of the glottis.

A mathematical expression was included in the proposed vibratory model to approximate the real displacement function of the glottis. This expression facilitates the study of vibratory patterns of an abnormal vocal fold and the investigation of the effects of vibratory patterns on voice quality where a non-sinusoidal displacement may be presented.

CHAPTER 3

VOCAL FOLD VIBRATION AND GLOTTOGRAPHIC WAVEFORMS

The vibratory model introduced in the last chapter provides a framework to describe the vibratory behavior of the human vocal folds. In this chapter we relate the vibratory motion of the vocal folds to glottographic signals frequently used to (1) validate the vibratory model with measurable signals and (2) develop the relation between glottographic waveforms and vibratory patterns of the vocal folds.

The first section of this chapter describes the derivation of the glottographic waveforms from the displacement of the glottis. In the second section we consider simulation results for glottographic waveforms and explores the relation between glottographic waveforms and the vibratory pattern of the vocal folds. The effects of laryngeal control on the projected glottal area and the EGG signal are explored in the third section . In the fourth section we demonstrate the influence of the vibratory pattern on the primary excitation source, the glottal volume velocity.

The Derivation of the Glottographic Waveforms

The three glottographic waveforms, EGG, PGG, and inverse-filtered glottal waveforms, have been related to

glottal characteristics, i.e., lateral contact area [Childers, Smith, and Moore, 1984], glottal area [Kitzing, 1985], and glottal volume velocity [Wong, Markel, and Gray, 1979], respectively. To establish the relation between the glottographic waveforms and the vibratory patterns of the vocal folds, we first express these glottal characteristics in terms of the glottal displacement (see Figure 3-1). Formulas are then applied to the internal glottal signals for simulation of glottographic waveforms. Comparisons between the synthetic glottographic waveforms and measured waveforms such as EGG signals or estimated waveforms such as those from the inverse filtering allows evaluation of the proposed vibratory model and the transduction formula.

The Derivation of Projected Glottal Area

The formulation of the projected glottal area follows the work by Titze [1984]. The projected glottal area is the area outlined by the glottis when seen from above the glottis. The glottal area is believed to be the primary descriptor of the glottal excitation. The glottal area becomes zero when the glottis is completely closed along its length.

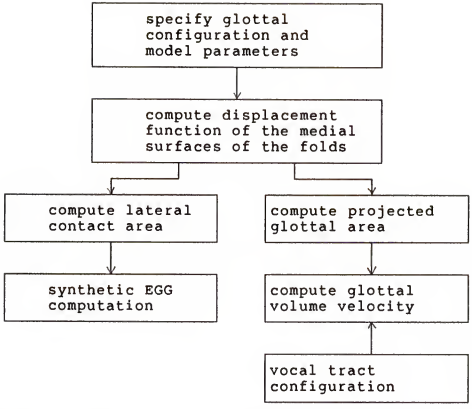
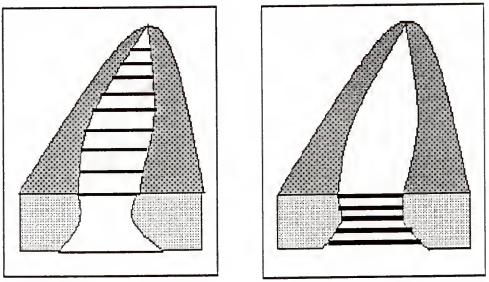


Figure 3-1 Flow chart for calculation of glottographic waveforms

Consider the length L of the glottis to be divided into a series of differential lengths dy and the 3-dimensional glottis to be sliced into a series of differential ducts (see Figure 3-2-a). The total projected glottal area is then the summation of the minimum lateral distances in each of the vertical ducts along the length of the glottis. This can be expressed as

$$A_g(t) = \int_0^L d_{min}(y,t)dy \qquad (3-1)$$



(a) (b)

Figure 3-2 Two methods for computation of the effective glottal area.

where $d_{min}(y,t)$ is the minimum positive value of glottal width in a differential vertical duct and dy is the increment along the length. In discrete form, the equation (3-1) becomes

$$A_g(n) = \sum_{i=1}^{M} d_{min}(i,n)dy \qquad (3-2)$$

for the nth time step and a total of M ducts along the length.

In practice these minimum glottal widths may occur on different horizontal planes. For the computation of the glottal volume velocity, it may be more appropriate to define the minimum glottal area as

$$A_g(n) = \min_{k=0,d} [A(k,t)] \qquad (3-3)$$

where A(k,t) is the kth glottal area, considering the 3-dimensional glottis to be sliced into a series of differential horizontal plates (see Figure 3-2-b). Simulation results show that there is no significant difference between results from equation (3-2) and equation (3-3).

The Derivation of Vocal Fold Contact Area

Vocal fold contact area is the lateral area of contact between the folds when they come together. To compute the total contact area, an infinitesimal amount of contact area dydz is added whenever the glottal width goes to zero at any coordinate (y,z) on the glottal mid-plane. The total contact area is then the summation of the partial contact areas along the length and the depth of the vocal fold or

$$A_c = \Sigma\Sigma c(y,z)dydz \qquad (3-4)$$

where $c(y,z) = 1$ for $d(y,z) \leq 0$

$= 0$ for $d(y,z) > 0$

For numerical computation of the double summation in equation (3-4), an interpolation algorithm may be needed to minimize the effect of quantized levels of contact dydz if the process of dividing up the sample area is crude.

In the above formulation, the vocal folds are allowed to overlap. We may consider the fold tissues to be incompressible and every movement of the vocal folds results in their deformation in another direction. During fold collision, the folds press against each other and cause a change in the thickness of the vocal folds. For a better approximation to the collision process, the thickness at the collision surface should be dynamically adjusted according to the degree of overlap (see Figure 3-3).

Our approach for calculating the varying thickness during the collision is to use the incompressibility of the fold tissue. The incompressibility property requires that the total volume of the vocal folds remains the same. This implies:

$$(x(t)-x_0)d_0 = (x_c(t)-x_0)d(t) \qquad (3-5)$$

where $x(t)$ is the lateral displacement of a unit area, x_0 the position of the boundary, d_0 the nominal unit contact area if there is no overlap, x_c the location where contact occurs, and $d(t)$ the adjusted contact area of the vocal folds. One immediate result of this approximation is that the EGG

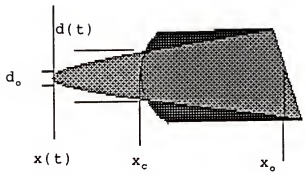


Figure 3-3 A frontal cross-section of the right fold demonstrating the idea for computation of the lateral contact area as folds collide. The shaded region represents the right fold as if there were no collision. The dark region represents the right fold that stops at the collision plane.

waveform in the vicinity of the EGG minimum would be rounded rather than flat. This result is a consequence of the change in contact area after the first collision.

The Simulation of EGG Waveforms

The electroglottograph is a device that measures electrical impedance in the vicinity of the larynx. It is generally accepted that the EGG reflects the changes in the lateral area of contact [Childers and Krishnamurthy, 1985; Kitzing, 1986]. Phonation alters the impedance, most likely due to changes in the current paths within the larynx. These changes occur when vocal fold motion alters the glottal configuration. A closed glottis would create a relatively low impedance compared to an open glottis because the current would have more direct current paths between electrodes. However, the vocal fold contact area has never been measured in vivo, only inferred. Titze [1988c] used excised canine larynges to examine the relation between EGG signals and the corresponding dynamic vocal fold contact area. His results did not refute the hypothesis of a linear relation between contact area and the EGG signal. Thus, it was assumed in this study that the change in electrical impedance measured by the EGG is inversely proportional to the changes in lateral contact area. The lateral contact area and EGG signal are then related as follows:

$$EGG(t) = \frac{k_1}{c(t) + k_2} \qquad (3\text{-}6)$$

where $c(t)$ is the contact area, k_2 represents the shunt impedance of the adjacent tissues, and k_1 is a scaling constant.

The Derivation of the Glottal Volume Velocity

The glottal area function is the major descriptor of the voicing excitation. Variations in glottal area determine many key characteristics of the glottal volume velocity, both in the time domain and the frequency domain. Research also shows that the interaction between the glottal source and the vocal tract plays an important role in determining the naturalness and the quality of synthesized speech [Fant, 1986]. The following major effects of source-tract interaction have been identified [Ananthapadmanabha and Fant, 1982]: (1) skewing, (2) truncation, (3) dispersion, and (4) superposition.

The glottal volume velocity is affected aerodynamically by changes in the subglottal pressure and the supraglottal configuration. The feedback effect of the vocal tract on the glottal volume velocity can be modeled by the circuit analog of a voice production system shown in Figure 3-4. The glottis acts as a flow-dependent acoustic impedance modulated by the time-varying glottal area [Fant, 1979b]. As seen from the glottis, the subglottal system can be characterized by a lung pressure source, P_L, in series with a single subglottal

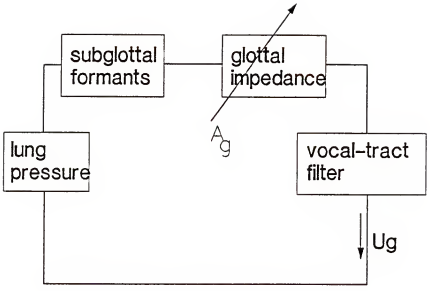


Figure 3-4 A circuit analog for interactions between the glottal excitation and the supraglottal structure.

parallel resonant impedance. The loading impedance of the vocal tract is modelled as a series of parallel RLC circuits.

Glottal impedance. Van den Berg [1968] proposed an empirical formula for the steady glottal flow based on static laryngeal models with uniform glottal widths. His formula for the transglottal pressure drop, ΔP, is:

$$\Delta P = (k\rho U^2)/(2A^2) + (12\psi DHU)/A^3 \qquad (3\text{-}7)$$

where ρ is the density of air, ψ is the coefficient of viscosity, U is the volume velocity, ΔP is the transglottal pressure, A is the glottal area, D is the glottal depth, and H is the length of the glottis. The first term in equation (3-7) is called the kinetic resistance drop and the second term the viscous drop. For a glottis with non-uniform widths, the second term is small and can be omitted. During phonation, the glottal area changes continually. Thus, under dynamic conditions the glottal impedance incorporates an additional term corresponding to the glottal inertance. Assuming a constant glottal depth, D, the glottal inertance is

$$L_g(t) = \rho D/A_g(t) \qquad (3\text{-}8)$$

The pressure drop across L_g is

$$\Delta P_L = L_g dU_g/dt + U_g dL_g/dt \qquad (3\text{-}9)$$

The complete equation for the flow without load is given by

$$K_k U^2 + R_v U + L_g dU/dt + U dL_g/dt = P_{t_g} \qquad (3\text{-}10)$$

where K_k and R_v are the kinetic resistance and the viscous resistance, respectively.

Load representation. Both the subglottal cavity and the vocal cavity are represented as lumped-parameter systems as seen from the glottis, i.e., cascade of parallel "RLC" tuned circuits. Each tuned circuit represents a subglottal or a supraglottal formant and values of R, L, and C are tuned according to the resonant frequency, bandwidth, and inductance associated with a formant. Ananthapabmanabha and Fant [1982] showed that the effects of the higher subglottal and supraglottal resonances upon the glottal volume velocity waveform were insignificant.

Simulation Results and Their Interpretations

General Observations of Glottographic Waveforms

While the glottographic signals (PGG, EGG, and inverse filtered glottal flow) are nonlinear functions of the complex vocal fold vibratory pattern, their shapes are relatively simple. This is because each waveform represents an average measure of the vibration. As an example, typical samples of the glottal area function and the EGG waveform are shown in Figure 3-5.

The Glottal Area. The configuration of a typical period of projected glottal area waveform is characterized by three segments: horizontal, rising, and falling segments, called closed phase, opening phase, and closing phase, respectively. The closed glottal phase begins at the point of complete glottal closure, which is the instant at which the glottal

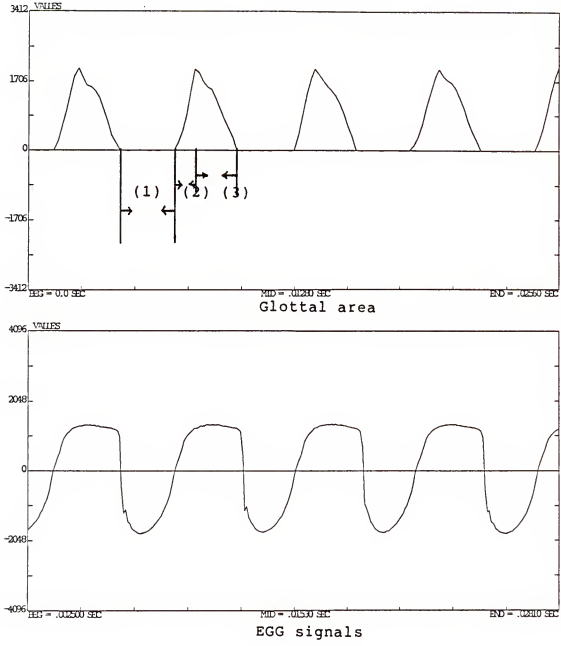


Figure 3-5 A short segment of measured EGG and glottal area waveforms.
1: closed phase 2: opening phase
3: closing phase

area curve decreases to zero. In some cases, there is no complete closure and hence the closed phase does not exist. The opening phase is the duration from the instant of glottal opening to the maximum value of the glottal area. The closing phase is defined as the time duration from the maximum glottal area to the minimum area. The sum of the opening phase and the closing phase is the open phase of the period. For the analysis of the glottographic waveforms, it has been common practice to calculate certain ratios, such as the open quotient (OQ) and the SQ. The OQ is defined as the ratio of the duration of the open phase to the pitch period. The SQ is defined as the ratio of the duration of the opening phase to duration of the closing phase and thus indicates the symmetry of the waveform. Studies have been conducted to correlate these quotients to changes in voice intensity and fundamental frequency [Hildebrand, 1976; Moore, 1976].

The EGG waveform. The EGG waveforms are usually described in relation to the glottal area. The closed phase, opening phase, and closing phase are also used to segment a period of the EGG waveform. It is somewhat difficult, however, to identify the exact instants of initial glottal opening and maximum opening from the EGG waveform. The EGG continues to decrease after the glottis closes. The minimum in the EGG, corresponding to the maximum lateral fold contact, occurs in the middle of the closed phase. The EGG begins to increase from its minimum while still in the closed phase. During the

opening phase, the EGG has two distinct phases. First, the EGG increases monotonically, reflecting the decreasing lateral contact between the folds as they start the opening phase. Once the folds have separated, the EGG remains constant while the folds pull further apart. At the beginning of the closing phase, the EGG continues to maintain a constant value while the vocal folds come together without contact. Then the EGG falls rapidly as a result of the folds coming into contact almost simultaneously from anterior to posterior.

The Effects of Model Parameters on Glottographic Waveforms

In the following simulations, a sinusoidal function is used as the time-dependent component in the displacement function. The term CONF refers to pre-phonatory configurations of the glottis and the maximum excursion profiles, specified by numbers assigned to pre-stored patterns in the simulator. The term PD stands for phase difference in the vertical direction and HPD stands for the phase difference in the horizontal direction.

To obtain more insight into the proposed model, in subsequent paragraphs we demonstrate the effects on the glottographic waveforms of varying one parameter at a time.

Effects of vertical time lag. The phase difference between the upper and the lower margins will affect the OQ and the peak values of the glottal area and the EGG signals. The behavior of the model for this factor is shown in Figure 3-6-

a. To investigate these effects, the model was set up with CONF = 2510 (the horizontal configuration #2, the horizontal maximum excursion profile #5, the vertical configuration #1, and the vertical maximum excursion profile #0) and DCAC = 0.6 (Figure 3-6-b). An increase in PD reduces the OQ and the peak value in the glottal area because the superior and the inferior portions of the vocal folds shade the glottis alternately. For the EGG signal, the asymmetry increases with the PD while the rising and the trailing slopes vary inversely with the PD.

The effect of horizontal time lags. Shown in Figure 3-7 is the effect of varying the horizontal phase difference from 0° to 60° where the model was set up with CONF = 2510, DCAC = 0.6, PD = 90°. An increase in HPD reduces the maximum of the glottal area slightly, but increases the OQ because the elements along the length of the glottis do not vibrate in phase. In this example, the increase in HPD makes the falling slope of the EGG signal more steep and, thus, results in an increase in asymmetry. In general, the effects of varying the HPD depend on the chosen model parameters. In other words, the relation between the temporal features and the HPD is not linear.

The effect of DCAC. Varying the DCAC will change the ratio of the abduction of the folds to the amplitude of vibration. The OQ of glottal area varies proportionally with the DCAC because the duration of the closed phase is reduced

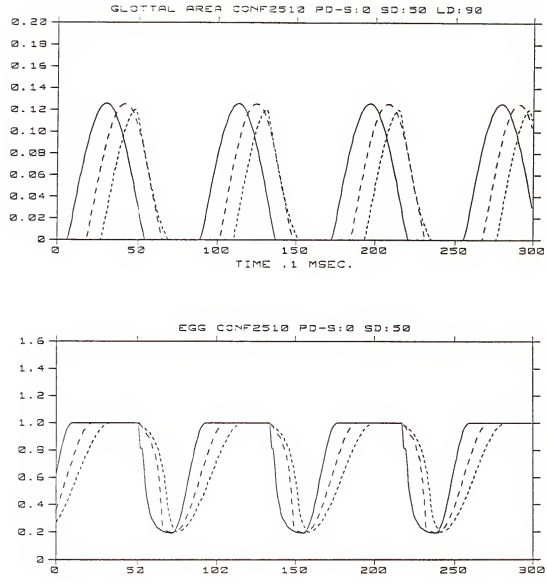


Figure 3-6-a Effects of vertical phasing on the simulated glottal area (top) and the EGG waveforms (bottom). Simulation conditions were: CONF = 2510, DCAC = 0.6, HPD = 0, and PD = 0 (solid line), PD = 50 (dashed line), or PD = 90 (dotted line).

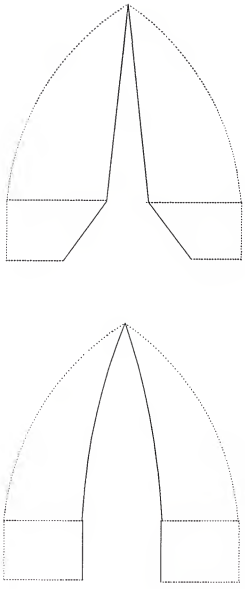


Figure 3-6-b The pre-phonatory configuration [2,1] (the upper graph) and the maximum excursion profile [5,0] (the lower graph). The notation [2,1] means the combination of the horizontal shape #2 and the vertical shape #1.

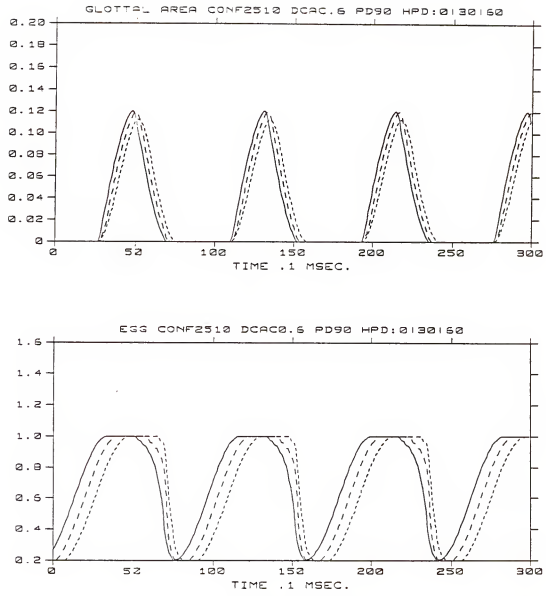


Figure 3-7 Effects of horizontal phasing on the simulated glottal area (top) and the EGG waveforms (bottom). Simulation conditions were: CONF = 2510, DCAC = 0.6, PD = 90, and HPD = 0 (solid line), 30 (dashed line), or 60 (dotted line).

when the vibrating amplitude is relatively small (Figure 3-8). The maximum contact (indicated by the minimum of EGG) varies inversely with the DCAC because some part of the folds never contact as the vibrating amplitude gets smaller (Figure 3-8). Note that the falling slope of the EGG waveforms get more steep as the closing proceeds from the lower to the upper edges. The vertical phasing causes this, as will be seen in the next section.

Varying the vertical amplitude profile. The behavior of the proposed model with respect to the vertical amplitude profile is shown in Figure 3-9-a where the model was set up with PD = 90°, HPD = 0°, and DCAC = 0.6. The three vertical amplitude profiles used for this run are shown in Figure 3-9-b. The glottal area appears symmetric when the superior and the inferior margins of the folds vibrate at the same amplitude. In contrast, it tends to skew to the left or to the right when the superior and the inferior margins vibrate at different amplitudes. The amplitude profile also affects the OQ and the peak value of the glottal area function. The EGG waveforms remain askew to the right for all three amplitude profiles, but the maximum contact area and the rising and trailing slopes vary.

Varying the amplitude profile along the length of folds. The three amplitude profiles shown in Figure 3-10-b were chosen to demonstrate their effects on the glottographic waveforms. The amplitude at the anterior end is kept near

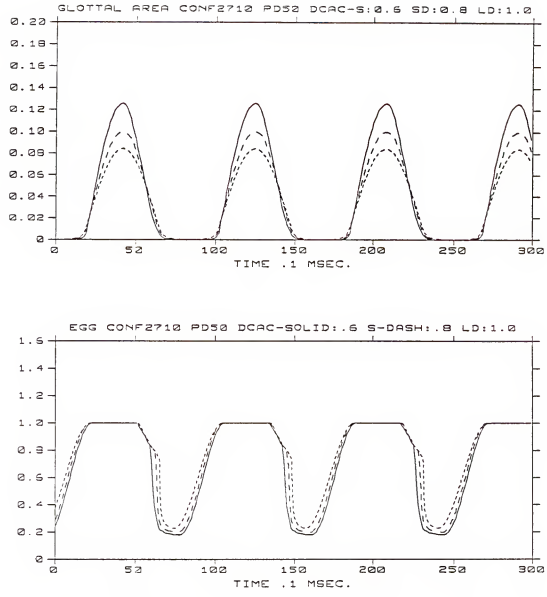


Figure 3-8 Effects of varying DCAC on the simulated glottal area (top) and the EGG waveforms (bottom). Simulation conditions were: CONF = 2710, PD = 50, HPD = 0, and DCAC = 0.6 (solid line), 0.8 (dashed line), or 1.0 (dotted line).

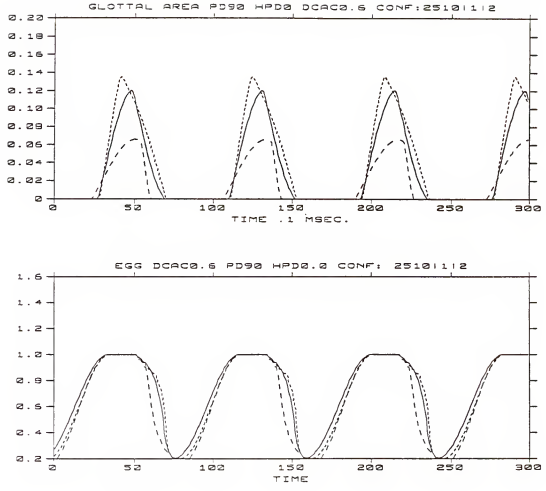


Figure 3-9-a Effects of varying the vertical amplitude profile on the glottal area (top) and the EGG waveforms (bottom). The simulation conditions were: PD = 90, HPD = 0, DCAC = .6, and CONF = 2510 (solid line), 2511 (dashed line), or 2512 (dotted line).

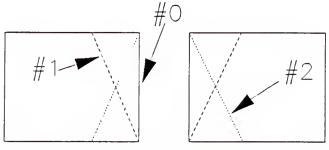


Figure 3-9-b A frontal cross-section of the glottis showing three vertical amplitude profiles: #0 (solid line), #1 (dashed line) and #2 (dotted line).

zero while the amplitude at the posterior end varies over a wide range. This variation mostly affects the region near the closing instant (Figure 3-10-a). It changes the abruptness at the closure and the OQ of the glottal area as well. Notice that zero amplitude at the posterior end results in an incomplete closure. This allows control over the glottal leakage.

Discussion

From the simulations above, we were able to determine that the proposed model can mimic the primary temporal features of the glottal area, the contact area, and the EGG waveforms to some extent. The model can also be adjusted to generate a great variety of waveshapes. The simulations performed give us a picture of how the glottal source parameters affect the glottographic waveforms. A summary of our results is given below.

1) The asymmetry in the lateral contact area waveform (skewing to the left) results from a combination of the convergent vertical shape of the glottis and vertical phasing. The vertical shape of the vocal folds is a time-varying function due to the vertical phasing. In our model the shape varies in form from an inverted cone to a funnel. Prior to closure, the shape becomes rectangular such that the contact area increases rapidly. On the other hand, the shape becomes trapezoidal prior to

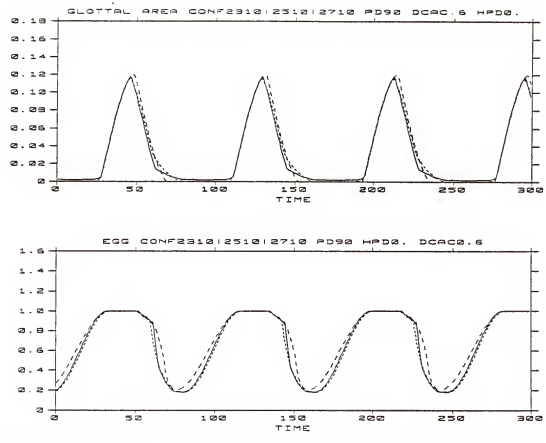


Figure 3-10-a Effects of varying the horizontal amplitude profile on the glottal area and the EGG waveforms. The simulation conditions were: PD = 90, HPD = 0, DCAC = .6, and CONF = 2310 (solid line), 2510 (dashed line), or 2710 (dotted line).

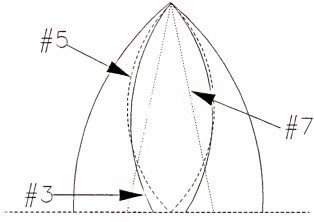


Figure 3-10-b A coronal cross-section of the glottis showing three horizontal amplitude profiles: #3 (solid line), #5 (dashed line) and #7 (dotted line).

opening such that the contact area decreases gradually. The degree of skewness depends on the vertical glottal shape and the phase difference. It is worth noting that this mechanism causes an asymmetry in the EGG waveform even when the corresponding glottal area function is symmetrical.

2) The asymmetry of the glottal area waveform is caused by different amplitudes of vibration for the superior and the inferior margins of the vocal folds as well as by vertical phase differences. This is illustrated in Figure 3-11. If the amplitude of excursion of the superior margins is larger than that of the inferior margins, then the resultant glottal area will skew to the left, otherwise it will skew to the right. This mechanism demonstrates that a skewed glottal area function is possible even when the time-dependent component of the displacement function of the vocal fold is symmetrical, as in our examples using a sinusoid function.

Simulation Examples

We shall now present some results for real glottographic waveforms. Our aim here is to determine if the proposed vibratory model is adequate to capture a variety of glottographic waveshapes by comparing synthetic waveforms against the real ones. The tasks are phonations of the vowel /i/ by three male speakers (AKK, DMK, and JMN) at two

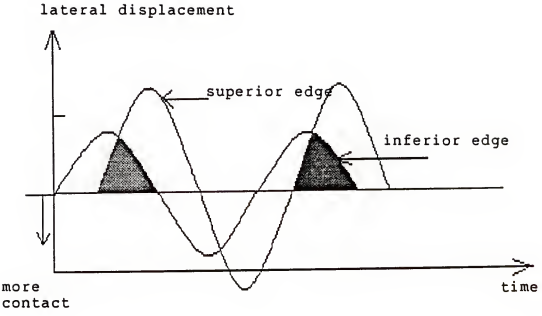


Figure 3-11 Illustration showing how the combination of vertical phasing and non-uniform vertical amplitude profile causes an asymmetry in the glottal area function. The dark region represents the projected glottal area.

different fundamental frequencies. In the following simulations, the dynamic component of the displacement function can be non-sinusoidal. Figures 3-12 to 3-15 are plots of measured glottal area, EGG signal, and the corresponding reconstructed waveforms for a number of typical tasks from our data base.

In Figure 3-12, the model was set up with CONF = 2510, DCAC = 0.8, PD = 90°, SQ = 1.2, and PQ = 1.5. The measured glottographic waveforms revealed a relatively adductive phonation, showing a large OQ. In Figure 3-13, the model was set up with CONF = 2512, DCAC = 0.5, PD = 90°, SQ = 1.0, and PQ = 1.5. The measured glottal area shows apparent asymmetry to the left; and the OQ is low. A large amplitude of vibration for the lower margin of the folds and a small DCAC were chosen to obtain the same asymmetry and OQ. The measured waveforms in Figure 3-14 show the typical characteristics of a high-pitch phonation, i.e. incomplete closure and large OQ. Here, the model was set up in the following conditions: CONF = 1000, DCAC = 1.2, PD = 30°, SQ = 0.8, and PQ = 1.5. Shown in Figure 3-15 are glottographic waveforms of an abductive phonation. Here, the model was set up in the following conditions: CONF = 1600, DCAC = 1.2, PD = 60°, SQ = 1.0, and PQ = 2.0.

The matching of model and measured waveform characteristics, (i.e., symmetry, OQ, slopes at discontinuous points, and overall waveshape), appears reasonable for all

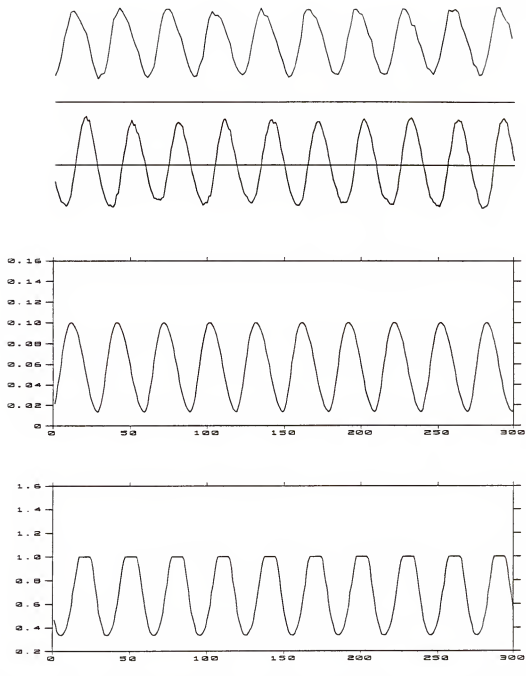


Figure 3-12 The graphs represent from top to bottom the measured glottal area, the measured EGG signal, the simulated glottal area, and the simulated EGG signal, respectively. The x-axis is the time axis in unit of 10^{-4} sec. The phonating task was an /i/ performed by a male.

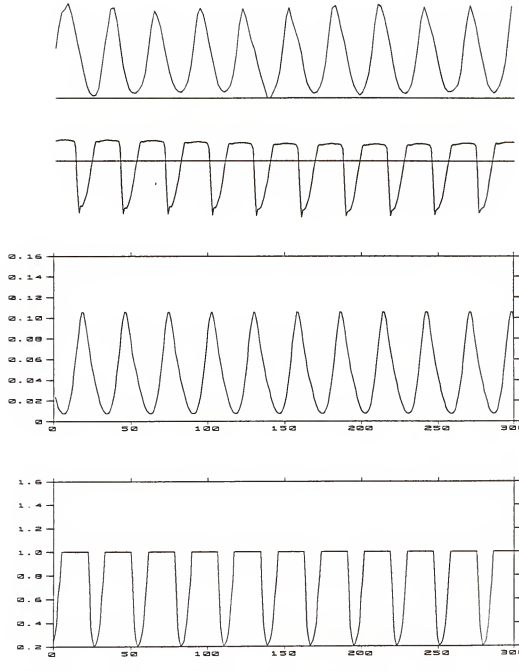


Figure 3-13 The graphs represent from top to bottom the measured glottal area, the measured EGG signal, the simulated glottal area, and the simulated EGG signal, respectively. The x-axis is time axis in unit of 10^{-4} sec. The phonating task was an /i/ by a male.

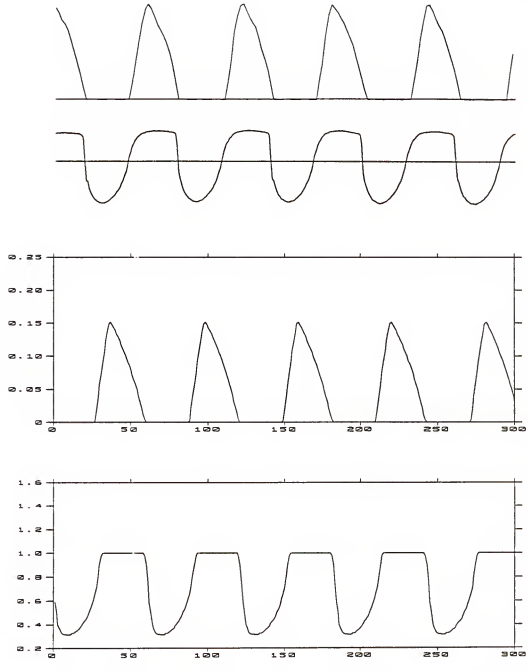


Figure 3-14 The graphs represent from top to bottom the measured glottal area, the measured EGG signal, the simulated glottal area, and the simulated EGG signal, respectively. The x-axis is time axis in unit of 10^{-4} sec. The phonating task was an /i/ by a male.

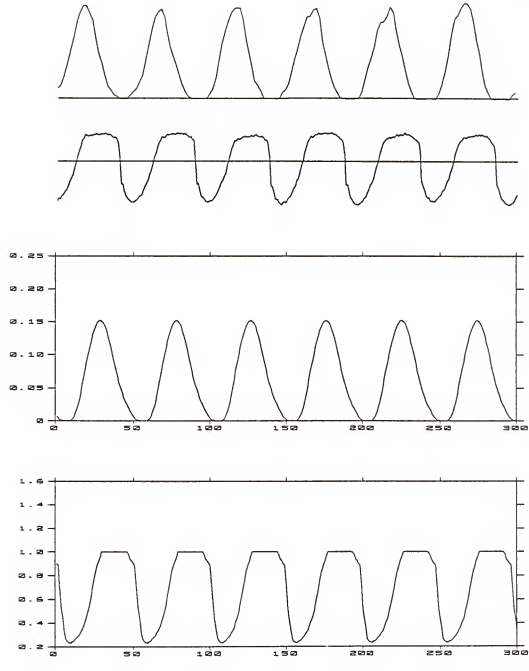


Figure 3-15 The graphs represent from top to bottom the measured glottal area, the measured EGG signal, the simulated glottal area, and the simulated EGG signal, respectively. The x-axis is time axis in unit of 10^{-4} sec. The phonating task was an /i/ by a male.

four tasks. There are, however, small discrepancies with regard to EGG signals. In Figure 3-14 and Figure 3-15, the measured EGG waveforms have a more rounded curve near the bottom of the valley; the simulated EGG signal has a flat top rather than a rounded one as in the measured EGG. An appropriate amount of longitudinal phase delay may be added to make the top of simulated EGG waveforms more rounded.

Glottographic Waveforms and Laryngeal Control

This section discusses the influence of laryngeal control on glottographic waveforms. This discussion is based on the simulation results of the two-mass model.

Intensity Control

We have shown in the previous chapter the behavior of the two-mass model under three subglottal pressure levels. The fundamental frequency, the vertical phase difference, and the amplitude of vibration all increase with the subglottal pressure. An increase in subglottal pressure also results in a skewness to the right and a more steep slope at closure in the displacement function (see Figure 2-5). Setting the kinematic parameters of the vibratory model according to these trends, we found that the SQ of the resulting glottal area and the falling slope of the EGG waveform shown in Figure 3-16 increase with subglottal pressure while the OQ of the

simulated glottal area varies inversely with subglottal pressure.

By comparison, a series of measured glottograms of three intensity levels at 161 Hz are shown in Figure 3-17. Note that for the low-intensity phonation the slope at the closure of the area function is less steep and that the SQ (0.88) of the area function is smaller. However, the OQs, which are 0.61, 0.67, and 0.55 for the phonations of 64, 70, and 74 db, respectively, do not vary inversely with the intensity as in the simulation. This result can be explained as follows. The two-mass model showed that subglottal pressure and laryngeal abduction may both affect the OQ. A possibility is that the subject might change laryngeal abduction across his phonations, while the degree of abduction remains unchanged in the simulation. Therefore, the assertion that the OQ increases inversely with the intensity is true only if laryngeal controls other than subglottal pressure do not mask this trend. In general, the slopes of area function or EGG waveform at closure are more consistent indications for intensity change, in spite of changes in other laryngeal parameters.

Pitch Control

We have shown in the previous chapter the behavior of the two-mass model under three levels of fold tension (0.8, 1.0, and 1.5). The fundamental frequency increases in proportion

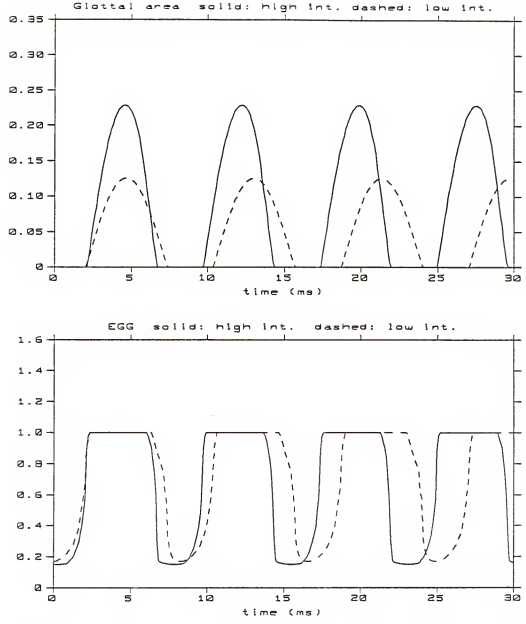


Figure 3-16 The simulated glottal area function (the upper graph) and EGG waveform (the lower graph) for phonations at two different subglottal pressures. The solid lines are for high pressure and the dashed lines are for low pressure.

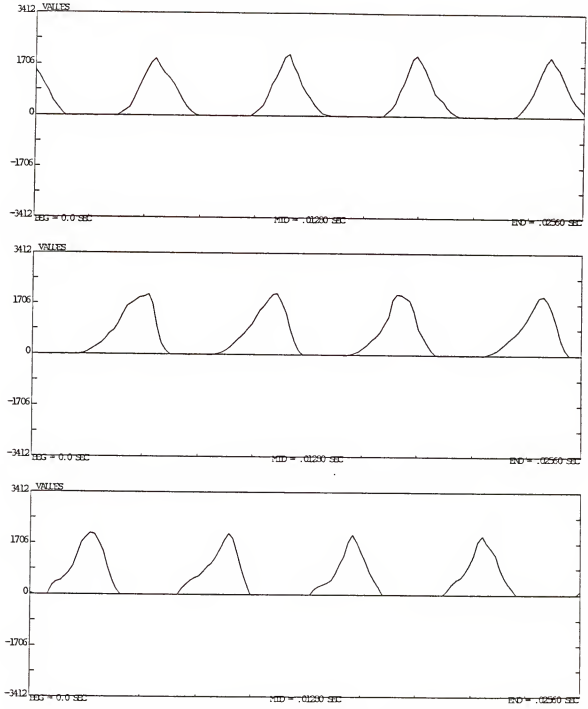


Figure 3-17-a The measured glottal area functions for phonations of three intensity levels at the fundamental frequency of 161 Hz. The intensities are 64, 70, and 74 dB, from top to bottom, respectively. The phonating tasks were an /i/ by a male.

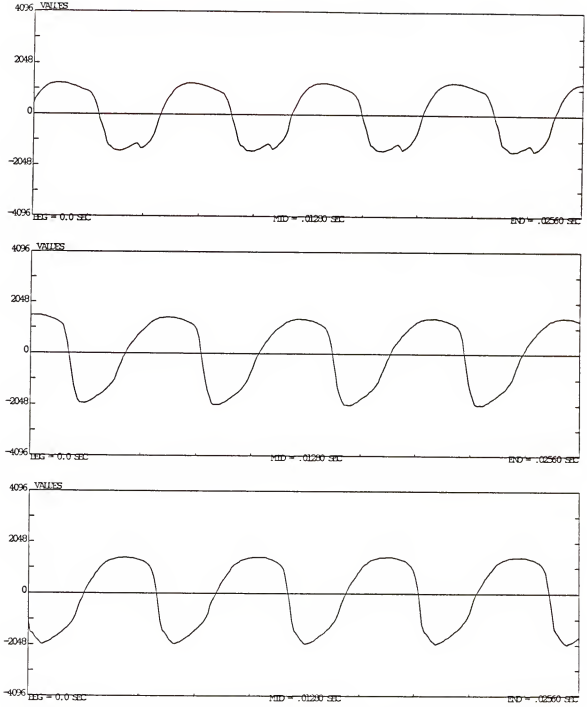


Figure 3-17-b The measured EGG waveforms for phonations of three intensity levels at the fundamental frequency of 161 Hz. The intensities are 64, 70, and 74 dB, from top to bottom, respectively. The phonating tasks were an /i/ by a male.

to the fold tension. The vertical phase difference and the amplitude of vibration vary inversely with the fold tension. The displacement function of the masses grows symmetrically and the amplitude of the upper mass outgrows that of the lower mass. Adjusting the vibratory model according to these trends, we found that the OQ of the simulated glottal area increases and the SQ decreases with the fold tension (Figure 3-18). The variation in skewness and OQ with fold tension can be verified by a series of measured area functions (Figure 3-19) of different pitches at the same level.

The Effects of Vocal Fold Configuration on Glottal Volume Velocity and Glottal Spectral Characteristics

We have described earlier in this chapter the procedure of deriving the glottal volume velocity from vocal fold dynamics, given the sub-glottal pressure and the vocal tract loading. In this section, the effects of vocal fold dynamics on glottal volume velocity, as well as its spectrum, are investigated since the auditory impression is directly related to spectral features [Ananthapadmanabha, 1984].

The Effects of the Source-Tract Interaction

The effects of the source-tract interaction on the glottal volume velocity can readily be seen in Figure 3-20. First, a skewing of the $U_g(t)$ to the right, (i.e. a slower rise and a faster decay than in glottal area), is observed. The

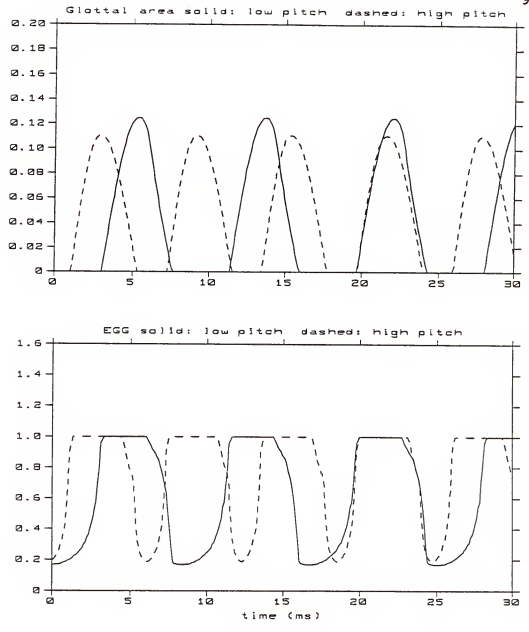


Figure 3-18 The simulated glottal area function (the upper graph) and EGG waveform (the lower graph) for phonations of two different pitches. The solid lines are for low-pitch phonation; the dashed lines are for high-pitch phonation.

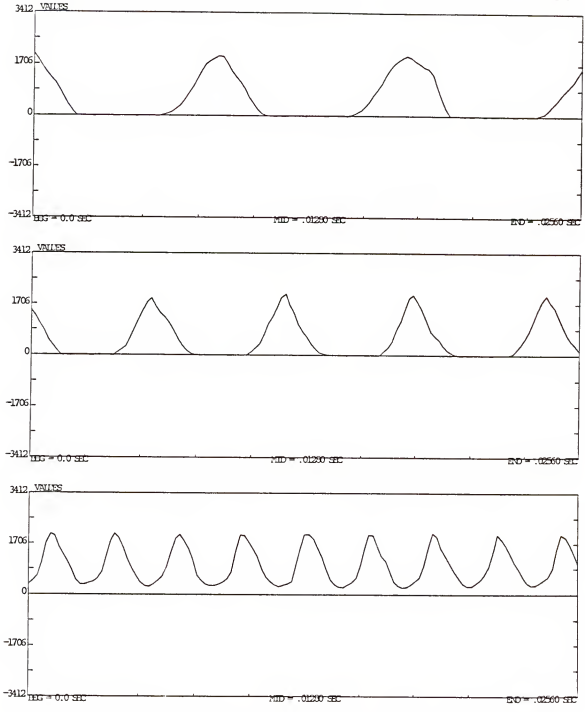


Figure 3-19-a The measured glottal area functions for phonations of three different pitches at the level of 70 dB. The pitches are 117, 168, and 336 Hz, from top to bottom, respectively. The phonating tasks were an /i/ by a male.

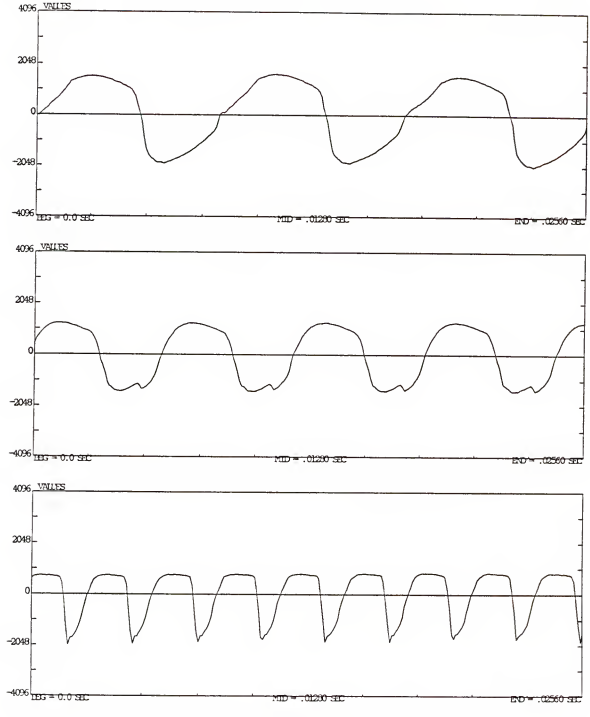


Figure 3-19-b The measured EGG waveforms for phonations of three different pitches at the level of 70 dB. The pitches are 117, 168, and 336 Hz, from top to bottom, respectively. The phonating tasks were an /i/ by a male.

skewing is caused by the overall vocal tract inertia [Rosenberg, 1981]. In addition, there is a "ripple component" on the rising portion of the glottal volume velocity representing the effect of a time-varying bandwidth and resonant frequency modulation due to the coupling over the open phase [Ananthapadmanabha and Fant, 1982].

To compare the effects of different vocal tract configurations on the waveshape of glottal volume velocity, three vowel configurations, /a/, /e/, and /i/, were used in the computation of the glottal volume velocity. The results are shown in Figure 3-21. The glottal flow waveform of vowel /i/ has the largest degree of asymmetry due to the large air inertia of this configuration. The glottal flow waveform of vowel /a/ has the largest ripple component due to the large coupling between the vocal tract and the subglottal system.

The Effects of Vibratory Pattern of the Vocal Folds

Several researchers have given a spectral description of the glottal volume velocity [Flanagan, 1958; Monsen and Engebretson, 1977; Fant, 1979a; and Sundberg and Gauffin, 1979]. In general, the glottal spectral features fall into three categories: (1) the general spectral slope, (2) the relative amplitude of the spectrum at fundamental frequency and its higher harmonics, and (3) the distribution of the zeros in the spectrum and their bandwidth. Ananthapadmanabha

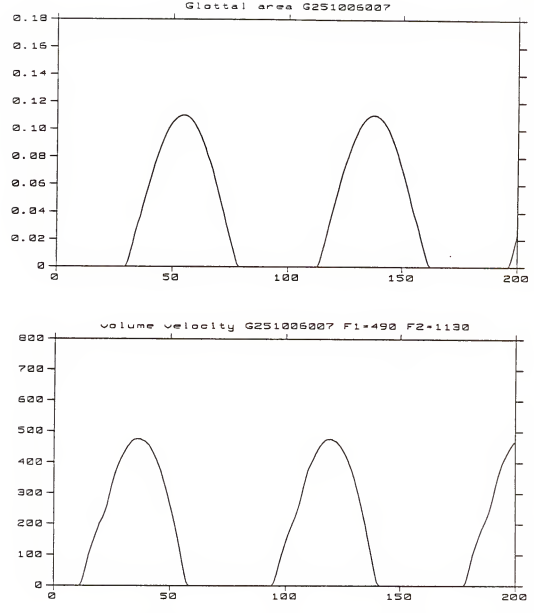


Figure 3-20 A segment of a typical glottal area function (upper graph) and the corresponding glottal volume velocity (lower graph). The x-axis is time axis in unit of 10^{-4} sec. The first and the second formants of the vocal tract configuration are located at 490 and 1130 Hz, respectively.

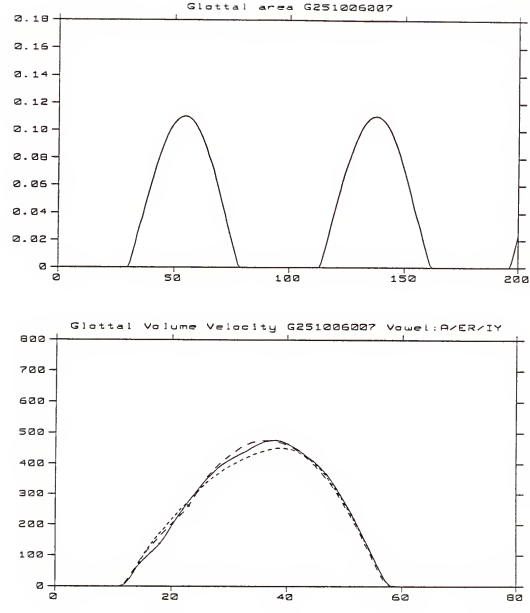


Figure 3-21 The effects of vowel configuration on the glottal volume velocity. The upper graph shows a segment of a typical glottal area function. The lower graph shows the resulting glottal volume velocity waveforms for vowels /a/ (solid line), /e/ (dashed line), and /i/ (dotted line).

[1984] identified the following spectral features shown in a typical glottal spectrum in Figure 3-22.

(1) Source spectrum peak (FL): This is the frequency at which the voice source pulse attains the maximum value.

(2) Source peak prominence (DL): This is measured as the dynamic range in dB between the spectral amplitudes at FL and the zero immediately following.

(3) First zero (ZL): This is the frequency of the first null after the source peak. This is an indirect measure of the effective open duration.

(4) Bandwidth of source peak (BL): Since the spectrum falls steeply after the first maximum, we have chosen the 12-dB-down bandwidth.

(5) High-frequency prominence (DH): The dynamic range in dB between the source peak and a reference high frequency gives a relative measure of the high-frequency dominance. This is also related to the spectrum roll-off (slope) in the dB/octave.

(6) Ripple content (DR): The dynamic range of the second spectrum peak is a measure of the ripple content of the glottal spectra. It also determines the relative excitation strengths at opening and closing instants. Larger DR results in deeper nulls, thus the DR determines the bandwidths of the nulls of the spectrum.

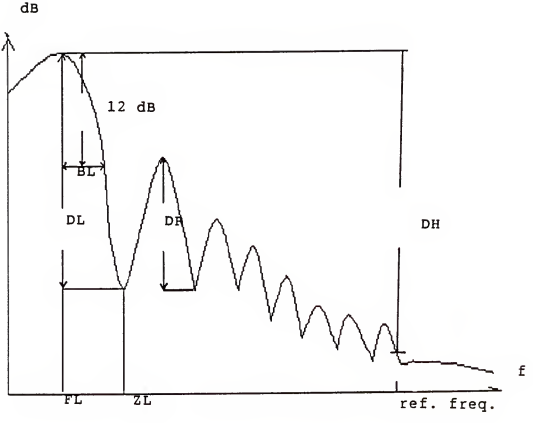


Figure 3-22 A hand-drawn spectrum illustrating the definitions of the spectral features.

In the following paragraphs, the effects of glottal adjustments on the spectra of the resulting differentiated glottal volume velocity are presented.

Effects of vertical time lag. To investigate the effects of vertical phase difference between the vocal fold's superior and the inferior margins of the vocal folds, the model was set with CONF = 2510, DCAC = 0.6, and HPD = 0°. The behavior of the vibratory model for three vertical time lags is shown in Figure 3-23. As the vertical phase lag increases, the ripple component becomes stronger and the source peak prominence becomes greater.

The effect of DCAC. Varying the DCAC will change the ratio of the abduction of the folds to the amplitude of vibration. The model is set up with CONF = 2510, PD = 90°, and HPD = 0°. The behavior of the vibratory model for three levels of glottal abduction is shown in Figure 3-24. An increase in the DCAC causes gradual changes in the glottal area and consequently an increase in the spectral tilt. An increase in DCAC also results in decreases in the source spectrum peak (FL), the first zero (ZL), and the bandwidth of the source peak (BL).

Varying the vertical amplitude profile. The spectra of the differentiated volume velocity with respect to the vertical amplitude profile are shown in Figure 3-25 where the model was set up so that PD = 90°, HPD = 0°, and DCAC = 0.6. The three vertical amplitude profiles used for this simulation

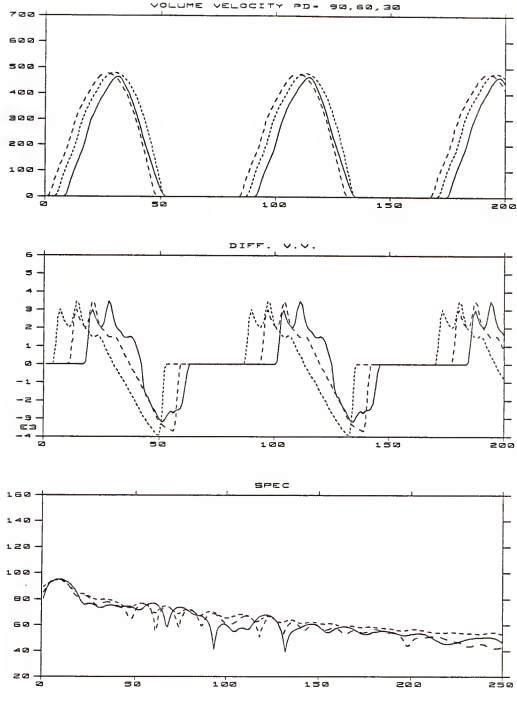


Figure 3-23 Effects of vertical phasing on the simulated glottal area (the upper graph), the differentiated glottal volume velocity waveform (the middle graph), and its spectrum (the lower graph). Simulation conditions were: CONF = 2510, DCAC = 0.6, HPD = 0, and PD = 0 (solid line), 50 (dashed line), or 90 (dotted line).

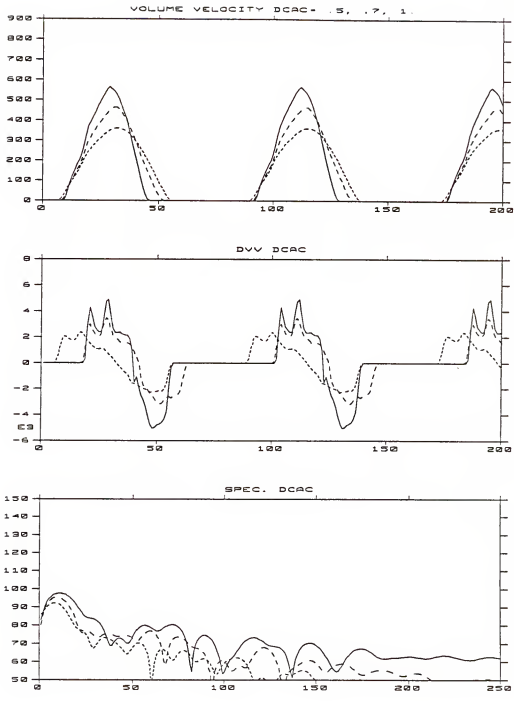


Figure 3-24 Effects of DCAC on the simulated glottal area (the upper graph), the differentiated glottal volume velocity waveform (the middle graph), and its spectrum (the lower graph). Simulation conditions were: CONF = 2510, PD = 90, HPD = 0, and DCAC = 0.5 (solid line), 0.7 (dashed line), or 1.0 (dotted line).

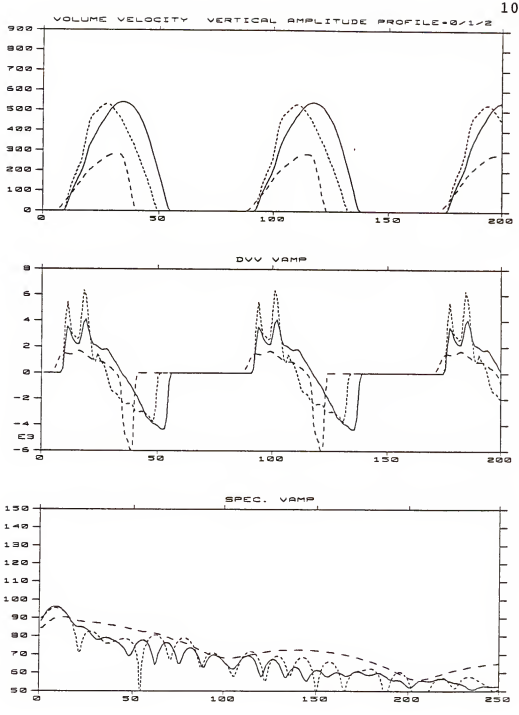


Figure 3-25 Effects of varying the vertical amplitude profile on the simulated glottal area (the upper graph), the differentiated glottal volume velocity waveform (the middle graph), and its spectrum (the lower graph). Simulation conditions were: DCAC = 0.6, HPD = 0, PD = 0, and CONF = 2510 (solid line), 2511 (dashed line), or 2512 (dotted line).

are shown in Figure 3-9-b. The spectrum derived from the right-skewing area function (dashed line) possesses a very special distribution of zeros. The high-frequency components are large because of the degree of skewness, while the zeros are widely spaced and very shallow. The spectrum derived from the left-skewing area (dotted line) has stronger ripple components because of the drastic changes on the rising slope of the corresponding volume velocity.

Summary

Based on the proposed vibratory model, several algorithms have been developed to simulate glottographic waveforms. To compute the area of lateral contact, the incompressible property of the fold tissue is used to adjust the thickness of the folds during the collision process. These simulations show how the vibratory patterns of the vocal folds affect the waveshapes of glottographic waveforms. In particular, mechanisms resulting in skewness in glottographic waveforms were identified and discussed.

Based on Fant's circuit for deriving the glottal volume velocity from the glottal area function, the proposed vibratory model is able to establish a relation between the glottal vibratory patterns and the glottal spectral features. The simulations demonstrate the effects of vertical phase difference, glottal abduction, and vertical amplitude profile on the shape of a glottal spectra.

CHAPTER 4

SIMULATION OF ABNORMAL VOCAL FOLD VIBRATION

The vibratory pattern of the vocal folds is a major factor relating laryngeal function to the sound produced. For example, breathy voices are usually associated with an incomplete glottal closure [Catford, 1977] and voice strains are often indicated by a very rapid glottal closure [Kitzing, 1983]. The first objective of this chapter is to investigate how the various normal and pathological modes of vibration come about as a result of various physical states of the larynx. The second objective is to find out how vibratory patterns of abnormal vocal folds manifest themselves on various measurable signals. In particular, the influences on two glottographic signals, EGG signals and PGG signals, are investigated.

In this chapter, the vibratory patterns of abnormal vocal folds are studied. Factors believed to contribute to abnormal vibratory patterns, such as imbalance of the vocal folds, irregular shape of the vocal folds, and abnormal tissue properties, are investigated and associated glottographic waveforms are simulated with the proposed vibratory model.

Previous Research

Perceptual Correlates

Hammarberg, Fritzell, Gauffin, and Sundberg [1980] performed perceptual analysis on utterances representing various voice disorders. They found five bipolar factors can describe the quality of most pathological voices. These factors are unstable-stable, breathy-harsh, hyper-hypofunctional, coarse-light, and head-chest register. The most important factor identified by factor analysis has been labeled "unstable-steady" because it includes the descriptive variables "bitonality, diplophonia, unstable pitch, jitter, grating, voice breaks, and unstable nasality" at one end of the continuum and the variable "restrained" at the other end. The second most important factor is "breathy-harsh", which includes the variables "breathy, wheezing, lack of timbre, and husky" at one end and the variables "creaky, and harsh" at the other end. These two factors accounted for 57.3% of the total variance in perceptual analysis for pathological voices. Unstable vibratory patterns and abnormal abduction of the glottis therefore are the two most common variables related to pathological voices.

Acoustic Correlates

The acoustic properties of abnormal voice have been investigated by many researchers. The acoustic analysis can be applied to radiated sound waveforms or indirect signals

such as glottal waveforms [Rosenberg, 1973] and residuals [Koike and Markel, 1975] derived by inverse filtering techniques. Acoustic measures reported to be associated with the detection of laryngeal diseases are:

(1) Perturbation in the fundamental pitch period and peak amplitude. These indices are often referred to as jitter and shimmer, respectively. Pitch and amplitude perturbation measures show differences between pathological and normal speakers.

(2) The noise level included in the voice signal. Inter-harmonic noise may originate in turbulent air flow due to incomplete glottal closure or irregular vocal fold vibration [Flanagan, 1958]. These spectral noise components are distributed over the spectrum in varying degrees, and the extent of the distribution depends on the severity of the disease. The presence of spectral noise contributes to "hoarseness," which is the first symptom of numerous voice pathologies.

(3) Cycle-to-cycle waveform variation. Irregular vocal fold vibration causes not only perturbation in pitch and amplitude but also variation in waveshape of adjacent cycles; and

(4) Average frequency-spectrum characteristics. Several spectrographic studies [Hiki, Imaizumi, and Hirano, 1976; Hammarberg, 1985; Muta, 1985] have shown that there are differences between the spectra of pathological voices and

the spectra of normal voices. The higher frequency harmonics of steady pathological vowels are greatly attenuated in comparison with their normal counterparts. The loss of high frequency harmonics may be caused by changes in the OQ or SQ. In particular, if there is no glottal closure, the higher harmonics are sharply attenuated.

Physiological Correlates

High-speed films of pathological vocal folds have revealed that, frequently, there are irregular vibration patterns [Moore, 1976]. Studies have been conducted to find the relation among tension imbalance, mass imbalance and irregularity of vocal fold vibration. Isshiki and Ishizaka [1976] showed with a bilaterally asymmetrical two-mass model that tension imbalance alone does not result in an irregular vibration of the vocal folds, but aggravates it. Titze [1988b] in his attempt to model asymmetrical conditions in the vocal fold tissues also pointed out that asymmetry either in mass or tension is not required for irregular vibrations to occur.

The vibratory pattern of a breathy voice is characterized by an incomplete and lax approximation of the vocal folds during phonation [Boone, 1971]. Because of low muscular effort, the lessened glottal resistance leads to a higher rate of airflow than in a normal voice. During breathy phonation,

the vocal folds do not come into complete contact, but instead simply "flap" in the airstream [Ladefoged, 1975].

One widely agreed physiological correlate of harshness is the laryngeal tension. Gray and Wise [1959] say that harshness results from overtensions in the throat and neck. Support for this line of reasoning comes from the fact that persons with harsh voices tend to initiate phonation with glottal attack [Zemlin, 1964].

Faulty Laryngeal Adjustments and Vibratory Pattern

Abnormal behavior of the human vocal folds may occur as a result of:

1. organic diseases, resulting from structural (anatomic) or physiological diseases of the larynx;
2. functional or psychogenic problems caused by faulty habits of laryngeal adjustments.

Two physiological phenomena are frequently seen in the vibration of abnormal vocal folds, incomplete glottal closure and irregular vibratory pattern. This contrasts with complete glottal closure in a pitch period and a nearly periodic vibration for most normal subjects. In this section, we are going to tune various physiological constants in the two-mass model of the vocal folds to investigate their effects on the vibratory patterns of the vocal folds. Of particular interest here are the elasticity (spring constant) and compliance (damping) of the vocal folds. The two-mass model discussed

in Chapter 2 is used here to study the behavior of the vocal folds during faulty laryngeal adjustments. Care must be maintained, however, in interpreting simulation results, as the two-mass model can only mimic a part of the vocal fold dynamics.

Incomplete Glottal Closure

Incomplete closure of the glottis during a pitch period is the main phonatory characteristic of a breathy voice. By comparison with modal voice, the mode of vibration of the vocal folds is inefficient, and is accompanied by slight audible friction. The mode is inefficient because the incomplete glottal closure results in a DC component in the air flow, which means less subglottal driving power is converted into sound energy.

An incomplete glottal closure alone does not make a voice sound breathy. Transglottal airflow and the waveshape of glottal flow also play a role in giving the perception of breathiness. The minimal glottal volume velocity for the production of a breathy voice is of the order of 90 to 100 cm^3/s and can be as high as 1000 cm^3/s [Catford, 1977]. A smoothly waveshaped glottal flow causes great attenuation of high-frequency harmonics so that friction noise in that frequency range becomes dominant and audible [Lee, 1988].

Underlying physiology

According to our simulations, an incomplete glottal closure during vibration can be the result of a relatively wide abduction, a large damping factor, or a large stiffness of the vocal folds. The following paragraphs explain how an incomplete glottal closure occurs as a result of different laryngeal settings.

Incomplete closure as a transient state. Vocal folds usually remain untouched during the build-up time of a vocal fold vibration. The build-up time required for the oscillation to reach a steady state increases with the abduction of the folds (see Figure 4-1). Therefore, the wider the abduction, the longer the incomplete closure lasts. There will be a complete glottal closure eventually, however, if the conditions are right. In other words, a wide abduction alone usually does not prevent vocal folds from complete closure but does delay its occurrence. There is a critical limit for the abduction beyond which the model does not oscillate under these conditions.

Incomplete closure as a result of increased stiffness. An increase in stiffness of the coupled spring K_c increases the build-up time required for the oscillation. For still larger values of K_c, close to the bounds of the oscillation range, the glottal area waveforms becomes sinusoidal on a dc component, and the glottis does not close completely. In contrast, an increase in K_2, with other conditions unchanged,

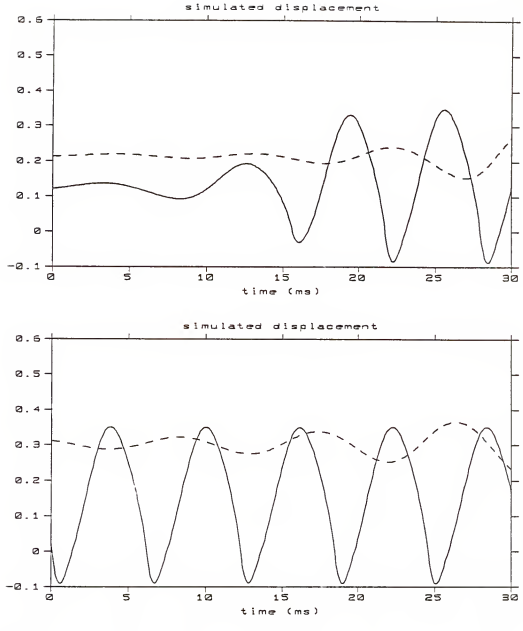


Figure 4-1 Effects of the abduction on duration of the incomplete closure. The upper graph shows that the build-up time increases with the abduction of the folds. The pre-phonatory abductions are 0.13 (solid line) and 0.22 (dashed line) cm respectively. The lower graph shows that the amplitude of vibration (dashed line) increases with time and the folds will eventually go across the sagittal line (come to contact).

decreases the vibration amplitude of the upper mass. This leads to incomplete closure of the upper mass, while the lower mass can close completely during the cycle. In our database, an incomplete glottal closure can often be seen in most high-pitch phonations where large fold tension is assumed (see Figure 4-2.). "Breathiness" is usually not perceived in such a situation because the transglottal airflow is not adequate to generate audible friction noise.

Incomplete closure as a result of increased damping. An increase in the damping factor of the masses results in an incomplete glottal closure and also suppresses the tendency to oscillate. Shown in Figure 4-3 is the effect of the damping factor on the displacement of the masses. The amplitude of vibration decreases, and no complete closure occurs as the damping factor increases. The large displacement amplitude increases the effective stiffness of the springs, and consequently increases fundamental frequency. Note that the displacement function also varies more gradually as the damping factor increases.

Incomplete glottal closure and glottographic waveform

The effects of an incomplete glottal closure on glottal area function are twofold: a DC component is added to the area function and the waveform is changed gradually. The effects of an incomplete glottal closure on the EGG waveform depends on how an incomplete closure occurs. If a glottal chink near

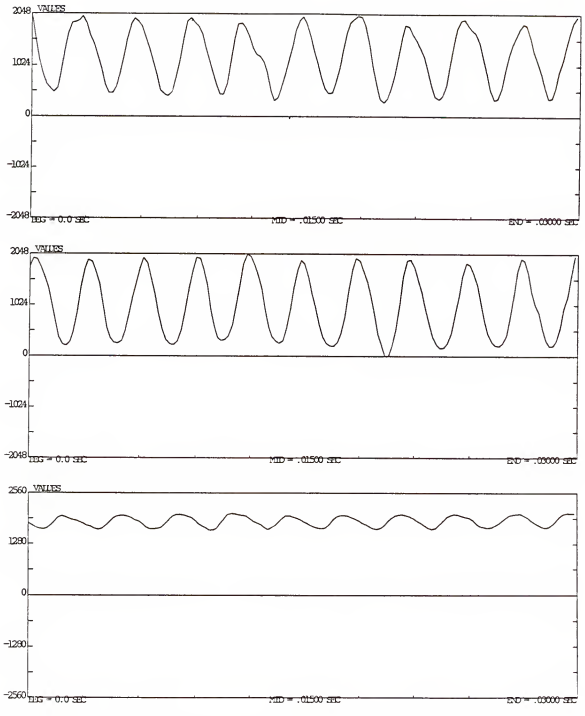


Figure 4-2 Three measured glottal area functions illustrate the correlation between an incomplete glottal closure and a high fold tension. The phonating tasks were an /i/ by three male subjects.

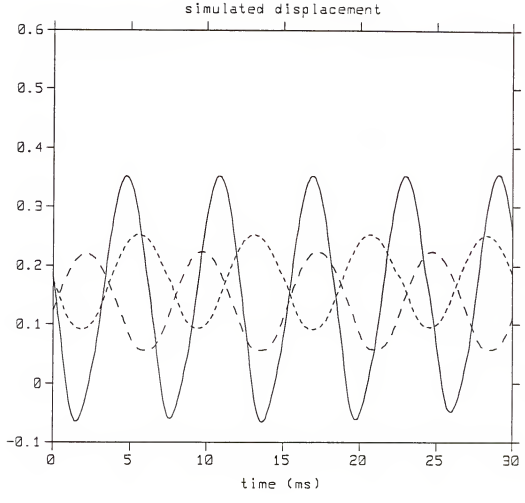


Figure 4-3 The effects of damping factor on the displacement of the glottis. The damping factors are 1.1 (solid line), 2.2 (dashed line), and 3.3 (dotted line), respectively.

the posterior portion of the glottis causes an escapage of air flow, then the EGG signal would appear normal (see Figure 4-4). If the degree of abduction is large and most parts of the vocal folds along the length of the glottis do not come in contact with each other, then the EGG signal shows a very short period of contact (see Figure 4-5.).

The influence of incomplete glottal closure on glottal spectra can be illustrated by comparing the speech spectrum of a modal phonation to the speech spectrum of a breathy phonation (Figure 4-6). The spectrum of breathy voice has a large low-frequency component and the frequencies between 1000 and 4000 Hz are attenuated abnormally. For frequencies above 4000 Hz, the normal and breathy spectra are similar. This may be caused by the friction noise in the breathy voice which boosts up the high frequency components and offsets the effect of the attenuation mentioned above in that frequency range.

Irregular Vibratory Patterns

Irregular vibratory patterns of vocal folds have been found frequently in speakers with voice disorders. Irregular vibratory patterns include perturbations in pitch and vibrational amplitude, and cycle-to-cycle variations in waveshape. For normal adults, the average variability of pitch from one period to the next appears to be less than 1 Hz [Monsen, 1978]. Perturbations in pitch can be further divided into two groups: subharmonics and random components.

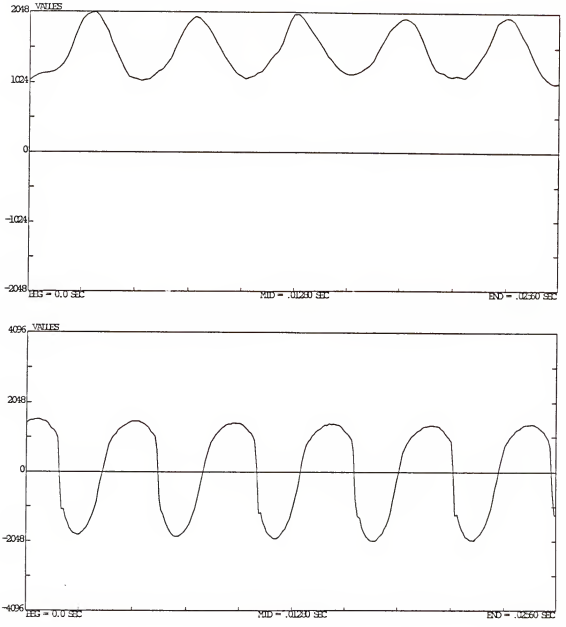


Figure 4-4 An example in which the glottal area function (the upper graph) shows no complete closure while the EGG signal (the lower graph) appears normal. The phonating task was an /i/ by a male.

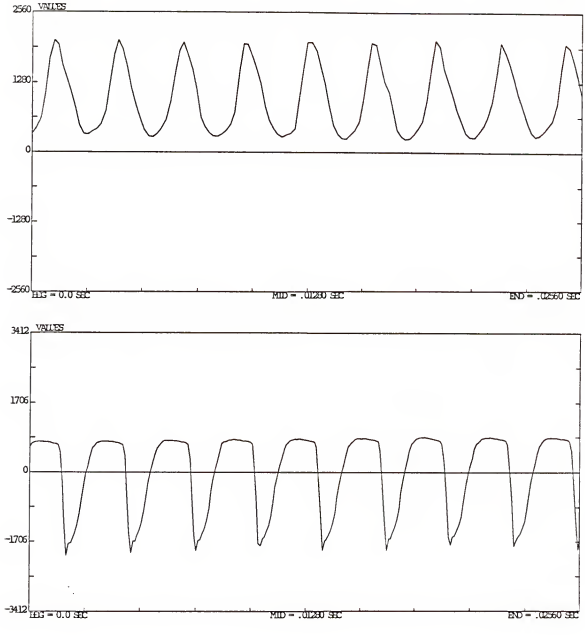


Figure 4-5 An example in which the glottal area function (the upper graph) shows no complete closure while the EGG signal (the lower graph) shows very short period of contact. The phonating task was an /i/ by a male.

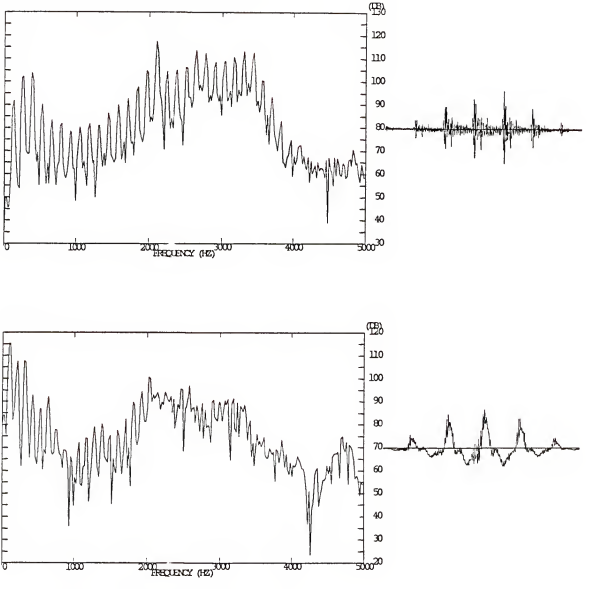


Figure 4-6 The speech spectra and speech waveforms of a modal phonation (the upper graph) and a breathy phonation (the lower graph). The phonating tasks were an /i/ by a male.

A 1/2 subharmonic (a waveform that repeats itself every other cycle) and an 1/3 subharmonic (a waveform that repeats itself every three cycles) can often be seen. If perturbation appears in an unpredicted way, we say there is a random component. A 1/2 subharmonic (which is diplophonic) is often found in a rough voice (see Figure 4-8), while random perturbation of pitch is usually associated with a hoarse voice (see Figure 4-10).

Underlying physiology

The underlying physiology of irregular vibration is not fully understood. Research into the mechanisms of an irregular vibratory motion of the vocal folds has been conducted in two directions. The irregular behavior of the vocal folds may be the result of a variation in laryngeal muscular control or it may be caused by abnormal biomechanical properties around the larynges, (e.g., extreme muscular tension or a difference in mass in the left and right folds).

By simultaneously recording EMG and speech signals, Baer [1979] found that changes in muscular tension contribute to pitch perturbation. Titze [1988b], in an attempt to model the irregular behavior of the vocal folds, used a multiple-mass discrete model of the vocal folds. The approach taken was to vary the mass, lateral stiffness, and longitudinal tension values, either locally or by changing the entire fold. These simulations showed that:

1) 1/2 subharmonics occur for a low value of longitudinal active tension regardless of the presence of any other changes, indicating that a lateral asymmetry in laryngeal tissue properties is not required.
2) Uniform unilateral increases in mass inhibit the subharmonics.
3) Local increases in mass enhance the subharmonics and produce a large random component.

Simulation of Irregular Vibratory Pattern

With the modified two-mass model, we are able to simulate some abnormal vibratory patterns by varying the following model parameters: masses, spring constants, and damping factors. Shown in Figure 4-7 are two simulated vibratory patterns with a 1/2 subharmonic. The upper graph shows a simulated displacement function under the following conditions: the abduction width is .03 cm; lung pressure is 9000 dynes/cm^2; the tension factor Q is 0.3; and the spring constant of the upper mass is one quarter of the typical value. In this simulation, the folds vibrate at a fundamental frequency of 60 Hz, in the range of vocal fry. A pitch period is composed of two pulses, with the first pulse aborting early, leaving a shallower dip between the two pulses. The resulting waveform has a deep valley and a shallow valley alternately presented.

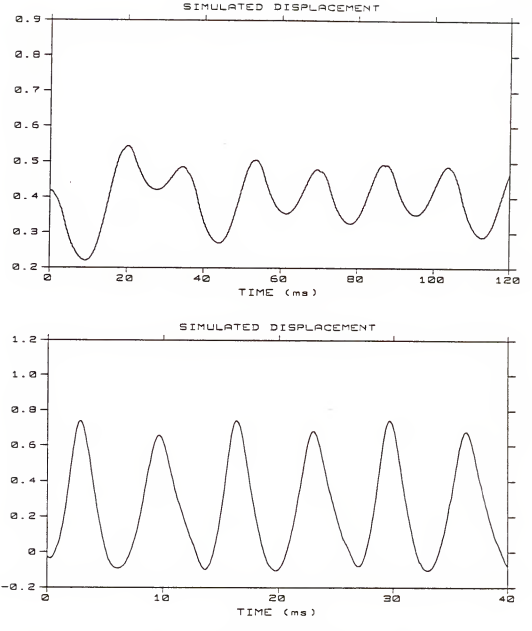


Figure 4-7 Two vibratory patterns that show 1/2 subharmonic. See related text for corresponding settings.

The lower graph shows simulated displacement under the following conditions: the abduction width is .03 cm, lung pressure is 15000 dynes/cm^2, the tension factor Q is 1, and the spring constant of the lower mass is one tenth of the typical value. The result shows a completely different pattern. Each pulse in a pitch period has a duration different from that of its adjacent neighbor.

In both simulations, at least one of the spring constants (k_1 in this case) should be kept low for the 1/2 subharmonic to appear. Simulations show that the appearance of jitter depends in a complicated way on abduction of the glottis, subglottal pressure, and tension of the vocal folds. The laterally symmetric model of the vocal folds used here is too restrictive for an irregular vibration to occur, i.e., only limited sets of parameter values would cause an irregular vibration of the vocal folds. In practice, however, patients with abnormal voices will keep showing their vocal characteristics in spite of possible laryngeal adjustments. Since lateral asymmetry can aggravate an irregular vibration [Titze, 1988b; Isshiki and Ishizaka, 1976], the introduction of various asymmetries into the model should allow irregular vibratory patterns to occur for a wider range of parameter values.

Relation to glottographic waveform

The irregularity of the vibration of the vocal folds can be seen in the resultant EGG waveform. Shown in Figures 4-8 and 4-9 are the speech and EGG waveforms of two rough phonations. One feature is common to the EGG signals-- 1/2 subharmonics.

The speech and EGG waveform of a hoarse phonation are shown in Figure 4-10. The amplitude and waveshape of the EGG signal change from cycle to cycle, but one consistent feature remains--sharp, narrow peaks in EGG signals. It agrees with the observations by others that a hoarse voice is often associated with excessive approximation of the vocal folds [Laver, 1980].

Simulation of Glottographic Waveforms for Laterally Asymmetrical Vocal Folds

Since disease and organic changes may not affect both vocal folds in the same way, it is necessary to postulate the influence of several factors when they are altered in only one of the folds.

Localized Protrusion

When a localized protrusion caused by edema or a tumor extends into the glottis, it will influence vibration variously in relation to its size, location, and firmness. Such an enlargement can prevent glottal closure, shorten the

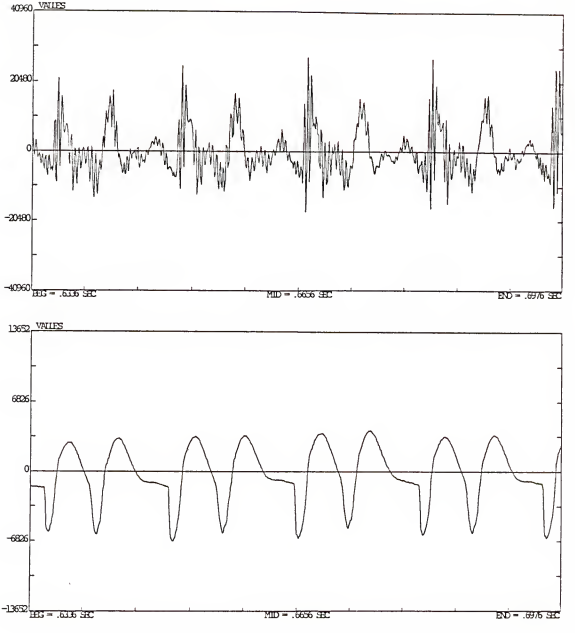


Figure 4-8 The speech waveform (the upper graph) and the EGG waveform (the lower graph) of a mimic rough voice. The phonating task was an /i/ by a male.

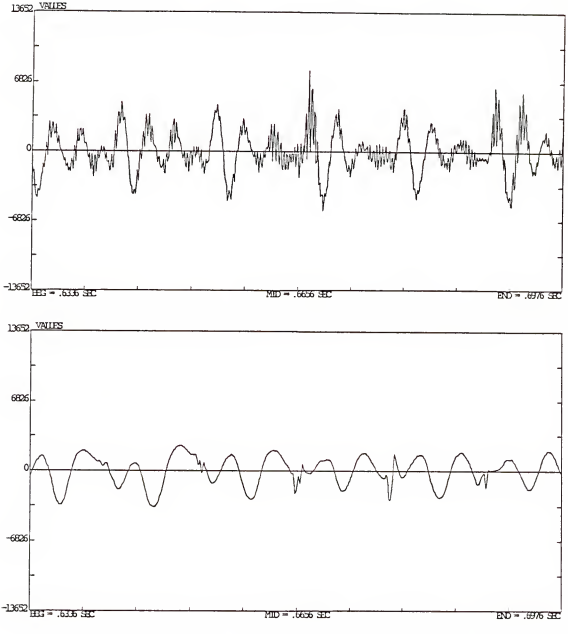


Figure 4-9 The speech waveform (the upper graph) and the EGG waveform (the lower graph) of a mimic rough voice. The phonating task was an /i/ by a male.

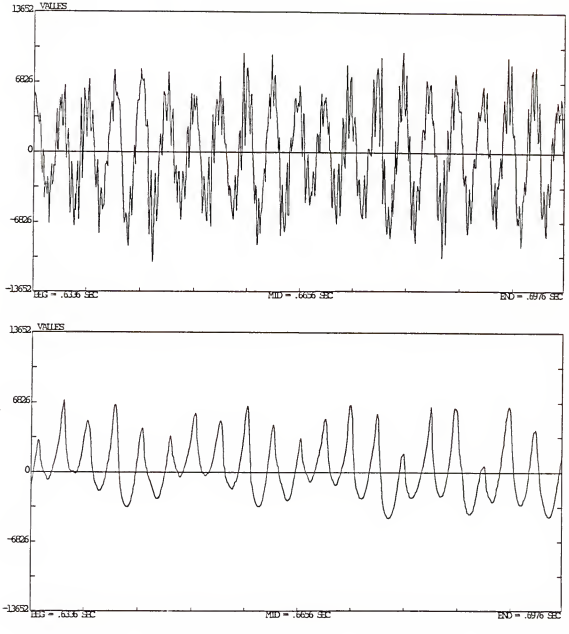


Figure 4-10 The speech waveform (the upper graph) and the EGG waveform (the lower graph) of a mimic hoarse phonation. The phonating task was an /i/ by a male.

vibrational portion of the fold, vary the mass in a circumscribed area, change the elasticity, or modify compliance. It is the size and location of the growth, not the type, that determines its effect on vibratory patterns [Moore, 1971].

It is expected that the EGG signal is sensitive to an irregularly shaped glottis because of its relation to the lateral contact area of the folds. Shown in Figure 4-11-a are a simulated EGG signal intended for reproducing the effect of a nodule and the EGG signals of a female patient who had a nodule on one of her vocal folds.

In the simulation, a nodule is added near the anterior portion of the glottis (see Figure 4-11-b). As the opening of the folds proceeds from the posterior towards the anterior end, the EGG progresses normally. But as the folds continue to separate beyond the location of the nodule, the EGG level remains constant. The EGG level rise again after the region near the nodule separates completely. The slope change of the simulated EGG in the opening phase is more abrupt than the change of the patient's data as the vibratory model fails to emulate the firmness of the nodule.

A similar change can be seen (Figure 4-12-a) as a nodule is added near the lower edge of the glottis (Figure 4-12-b). As folds initiate the closing phase from the lower edge, the EGG level falls more rapidly than if there were no nodule.

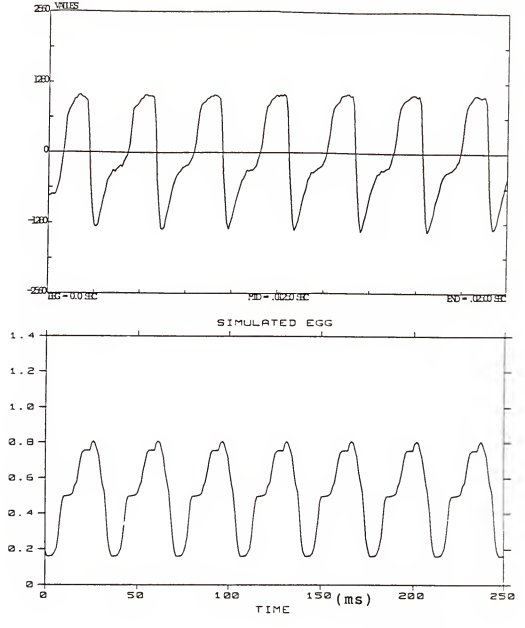


Figure 4-11-a The simulated EGG signal (the lower graph) intended for reproducing the effect of a nodule and the EGG signals (the upper graph) of a female patient who had a nodule on one of her vocal folds. The phonating task was an /i/.

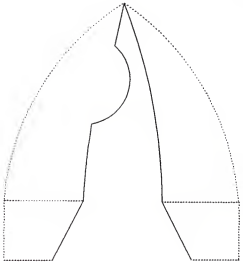


Figure 4-11-b A longitudinal equilibrium configuration used for simulation of nodule.

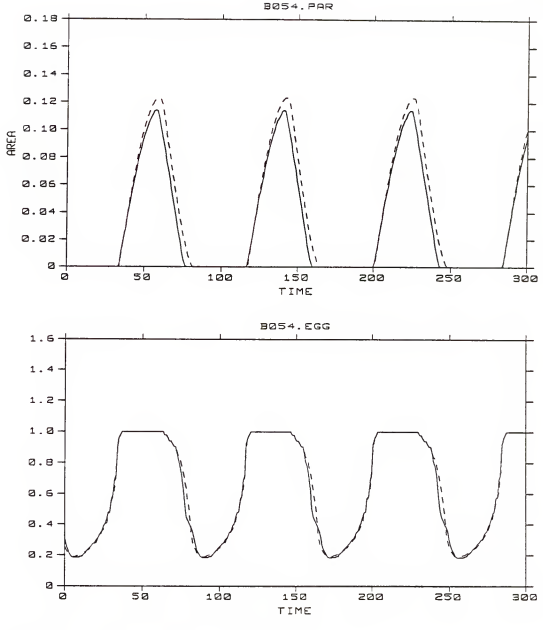


Figure 4-12-a The simulated glottal area (the upper graph) and the simulated EGG waveform (the lower graph) intended for reproducing the effects of a nodule near the lower edge of the glottis. The solid line: with a nodule; the dashed line: without any nodule. The time scales are in unit of 10^{-4} sec.

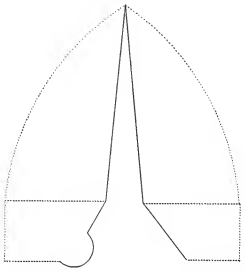


Figure 4-12-b A vertical equilibrium configuration used for simulation of nodule.

There is an abrupt change in EGG level after the separation of the folds passes the region near the protrusion.

Tension Imbalance

Isshiki and Ishizaka [1976] assigned different tension factors for the right and the left folds while keeping other parameters constant. Regardless of the tension imbalance, both vocal folds vibrate synchronously, but with differences in phase and maximum excursion of the left and right vocal-fold vibration. The tense fold side usually leads the lax side in movement.

Another interesting behavior is the dependence of bilateral amplitude of vibration on different degrees of tension imbalance. The amplitude of the lax side is smaller than that of the tense side with a tension imbalance. This change in amplitude of vibration causes a shift of the collision plane to the lax side.

To simulate the vibratory behavior of the vocal folds with a tension imbalance, the proposed vibratory model incorporated a lag between the motion of the left fold and the right fold. If the phase difference between folds is small, then there is no apparent change in the waveshape of area or EGG signal. There is an abrupt change in the falling slope of EGG waveform, however, if the phase difference is greater than 90 degrees (Figure 4-13). An abrupt change in the falling slope of area function is also observed if the displacement function

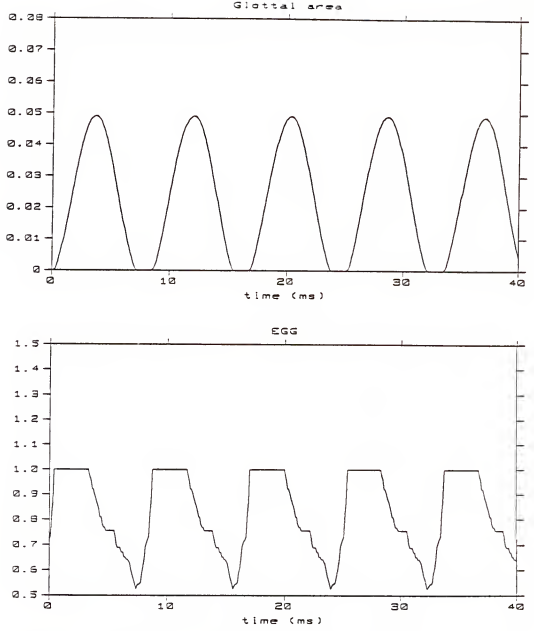


Figure 4-13 Effects of tension imbalance on the glottal area function (the upper graph) and the EGG waveform (the lower graph). Simulation conditions were: CONF = 2510, PD = 30, HPD = 0, SQ = 1, PQ = 2, DCAC = 0.6, and phase difference between folds = 120.

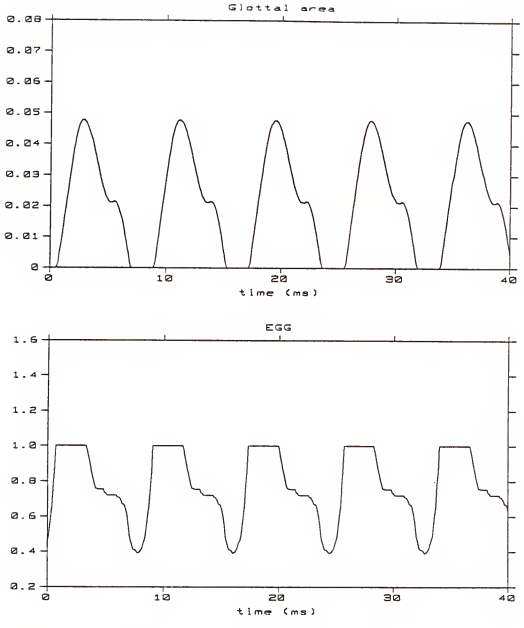


Figure 4-14 Effects of tension imbalance on the glottal area function (the upper graph) and the EGG waveform (the lower graph). Simulation conditions were: CONF = 2510, PD = 30, HPD = 0, SQ = 1.5, PQ = 2, DCAC = 0.6, and phase difference between folds = 120.

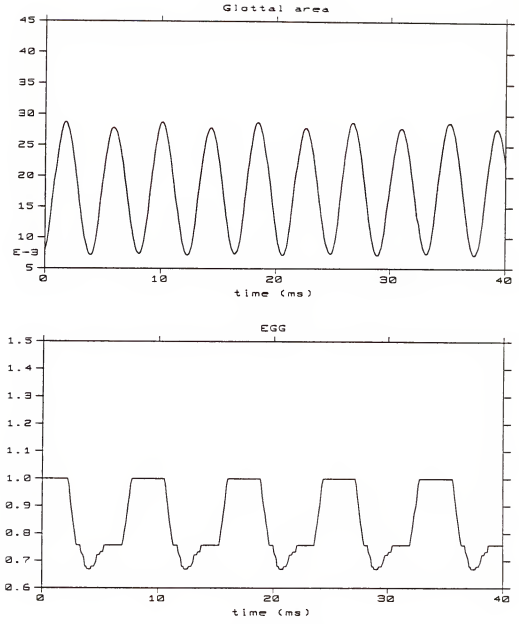


Figure 4-15 Effects of tension imbalance on the glottal area function (the upper graph) and the EGG waveform (the lower graph). Simulation conditions were: CONF = 2510, PD = 30, HPD = 0, SQ = 1.5, PQ = 2, DCAC = 0.6, and phase difference between folds = 180.

of the glottis is skewed and the phase difference is greater than 90 degrees (Figure 4-14). Note the appearance of an extra peak in the area function of a pitch period when the phase difference is 180 degrees (Figure 4-15). This is caused by the two folds alternately modulating the projected area when they are completely out of phase.

Unilateral Paralysis

"Unilateral laryngeal paralysis" means that the muscles attached to the arytenoid cartilage on the affected side cannot contract and, consequently, the patient is unable to close the glottis [Moore, 1971]. This positioning allows the breath to escape in greater amounts during phonation and results in a weak, breathy sound. The paralyzed fold may be fixed in any position between the median sagittal plane and maximum abduction, and is usually elevated slightly in relation to its healthy mate.

To simulate the glottographic waveforms when one of the folds is paralyzed, we choose an asymmetric maximum excursion profile in which the left fold does not vibrate and remains fixed. The resulting glottographic waveforms are shown in Figure 4-16. The glottal area function shows incomplete closure and becomes more gradual near the valleys. The EGG waveform also has abnormally long rising and falling slopes. The reason for the slowly-varying waveforms is that only one

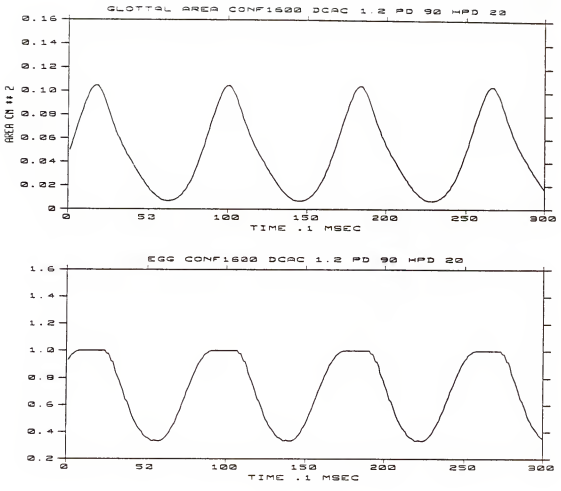


Figure 4-16 Simulated unilateral laryngeal paralysis. Simulated glottal area (top) and EGG waveform (bottom) for the following conditions: CONF =1810, DCAC = 1.2, PD = 90, and HPD = 20.

fold, instead of two, is vibrating. Consequently, the relative velocity between folds is reduced by one half.

Summary

We presented in this chapter the underlying mechanisms for two frequently observed abnormal vibratory patterns: incomplete glottal closure and perturbations in pitch and vibrational amplitude. The proposed model is able to predict the glottographic waveform as a result of these vibratory modes. Simulations show that perturbation in pitch and vibrational amplitude can occur even though parameters in a two-mass model are kept constant and are symmetrical for left and right folds.

The vibratory patterns for localized protrusion, tension imbalance, and unilateral paralysis have been simulated with the proposed model. The simulation shows that the EGG signals are more sensitive to an irregularly shaped glottis. A tension difference between the two folds can cause an abrupt change in the falling slope of the glottal area and the EGG waveform. Unilateral paralysis of the vocal folds can result in more gradual slopes at the opening and closing in both glottographic waveforms.

CHAPTER 5

VOICE QUALITIES AND THEIR ACOUSTIC AND PHYSIOLOGICAL CORRELATES

The purpose of this chapter is to investigate the relation between vocal fold vibratory characteristics and the resultant voice quality. The quality of voice is usually referred to as the total auditory impression the listener experiences upon hearing the voice of another speaker. The speaker's vocal quality is determined by his or her sex, health, age, emotional status, and phonation type. In this chapter, we confine our discussion to three major phonation types on the fundamental frequency continuum: modal voice, vocal fry, and falsetto. These different types of phonation are related to the production of the voice pitch and the pitch range of an individual. Although other factors have been used to describe register divisions, researchers [Boone, 1971; Hollien, 1974] generally agree that a particular vocal register is characterized by a certain pattern of vocal fold vibration, with the vocal fold approximated in a similar way throughout a particular pitch range. The following section describes the perceptual, acoustic, physiological, and aerodynamic aspects of the three registers.

Characteristics of the Vocal Registers

Modal Register

The modal register is so named because it includes the range of fundamental frequency that is normally used in speaking and singing [Hollien, 1974]. In a modal register, the vocal folds appear thick and rounded in the frontal projection. By means of lateral X-ray techniques, Damste [1969] observed that the length of the vocal folds systematically increases as the fundamental frequency of phonation increases. By using an X-ray laminagraph, Hollien and Colton [1969] observed that the thickness of the vocal folds systematically decreases as the fundamental frequency is increased. The vocal fold vibratory pattern of a modal phonation is characterized by a moderate fundamental frequency, wide lateral excursions, and a complete closure of the glottis during about one-third of the entire pitch period [Hollien, 1974].

Vocal Fry

Vocal fry voice is also called creaky voice or glottal fry voice. It occupies the lowest range of the phonation along the fundamental frequency continuum. The auditory effect of vocal fry is similar to a rapid series of taps, like a stick being run along a railing [Catford, 1964].

In vocal fry register, the arytenoid cartilages are tightly pressed together, so that the vocal folds can vibrate

only at the anterior portion [Ladefoged, 1975]. Using high-speed cinematography, Moore and von Leden [1958] and Timcke, von Leden, and Moore [1958] observed a double-pulse vibratory pattern of the vocal folds followed by a prolonged period of closed phase during vocal fry phonation. Also by using high-speed cinematography, Whitehead, R.L., Mets, and Whitehead, B.H. [1984] reported that single, double, and triple opening/closing patterns are all possible for vocal fry phonation.

Wendahl [1963] have suggested that "the primary criterion which must be met for a voice to be perceived as vocal fry is that the vocal tract be highly damped between glottal excitations." Coleman [1963] further established that vocal fry is perceived when the sound wave decays by approximately 43 dB between excitation pulses. A typical speech waveform of the vocal fry of the vowel /i/ is shown in Figure 5-1, along with the waveform of a modal phonation of the same vowel by the same subject. This figure reveals that rapid decay is characteristic of the vocal fry waveform. Hollien and Wendahl [1968] suggested that the perception of vocal fry is dependent on a low fundamental frequency and nearly complete decay of an acoustic signal, and not on the opening/closing gestures within a glottal cycle.

The effects of the vibratory pattern of a vocal fry register can be seen on the resultant speech spectrum, shown in Figure 5-2, along with the spectrum of a modal phonation.

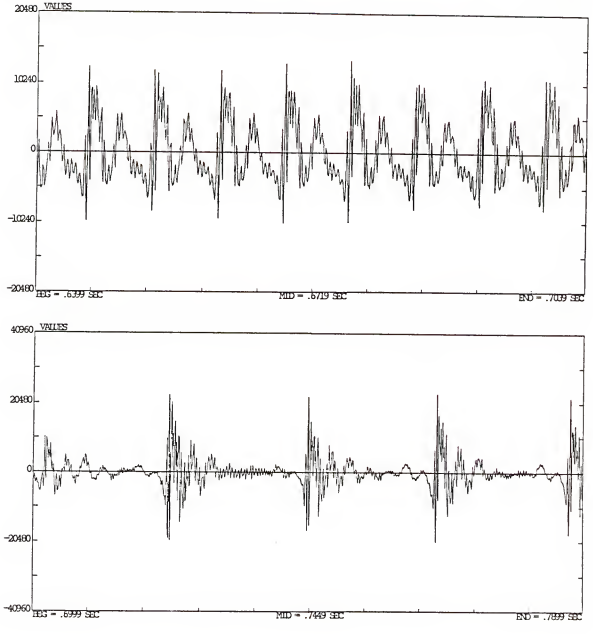


Figure 5-1 The speech waveforms of a typical modal phonation (the upper graph) and a typical vocal fry phonation (the lower graph). The phonating tasks were an /i/ by a male subject.

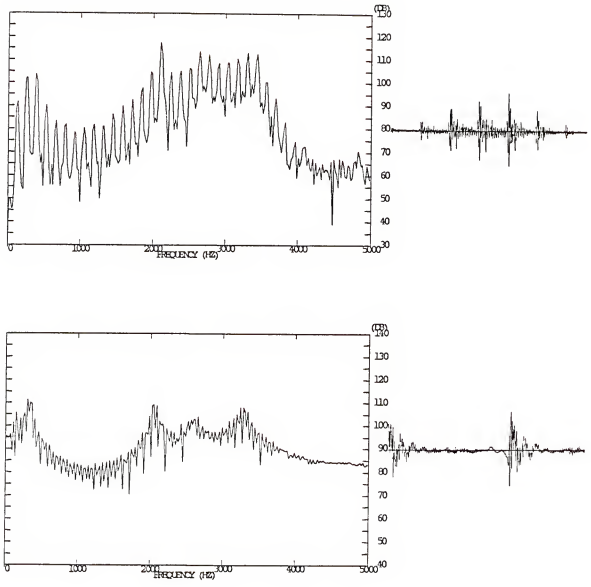


Figure 5-2 The speech spectra and the speech waveforms of a typical modal phonation (the upper graph) and a typical vocal fry phonation (the lower graph). The phonating tasks were an /i/ by a male.

First, note the clear formant structure caused by the low fundamental frequency. Secondly, observe that the spectral level for frequency bands above 3500 Hz is much higher than in modal phonation. This may be a result of a rapid glottal closure in vocal fry phonation.

Hollien, Damste, and Murry [1969] suggested that the control of fundamental frequency in vocal fry is not achieved by the same mechanism as in modal voice. In modal voice vocal fold length increases with fundamental frequency and vocal fold thickness is related inversely to the fundamental frequency. In contrast, for vocal fry, neither the length nor the thickness of the vocal folds seems to vary with changes in pitch. This suggests that muscle tension is not used for control of fundamental frequency. According to the two-mass model, adjusting fundamental frequency can be achieved through either tension control or subglottal pressure control. Since muscles function actively in only one way (they shorten by contraction), there is a limitation on the minimal tension that can be achieved. Due to the lack of a mechanism which would further reduce muscle tension to the range needed to produce such a low frequency vibration, control of fundamental frequency is managed by the aerodynamic component of the aerodynamic-myoelastic phonatory action, rather than the myoelastic component. The fundamental frequency of vocal fry ranges form 30 to 90 Hz.

Falsetto

The falsetto register occupies the higher portion of the fundamental frequency continuum. In a falsetto register, the vocal folds are stretched in the longitudinal direction, and the vertical cross-section of the edges of the vocal folds are extremely thin [Hollien, Coleman, and Moore, 1968]. The posterior cartilaginous portion is joined tightly so that there is little or no posterior vibration. The vibratory excursion in the lateral direction is also restricted by the stretched folds. The vocal fold vibratory pattern of the falsetto register is characterized by a high fundamental frequency, small lateral excursions, and momentary or even missing contact between the vocal folds [Hollien, 1974].

The more gradual opening and closing of the glottis in a falsetto phonation strongly affects the shape and slope of its spectrum. For comparison, the speech spectrum of a typical falsetto phonation is shown in Figure 5-3 along with the spectrum of a modal phonation by the same speaker. Higher harmonics are greatly attenuated due to the fact that the slope of the source spectrum for falsetto phonation is much steeper than in modal voice, falling at about -20 dB per octave. Also, whereas the speech spectrum of modal voice falls off more steeply with increasing frequency, falsetto seems to have more constant decrement. The harmonics for falsetto are widely separated in frequency, resulting in a "thin" voice quality [Zemlin, 1964].

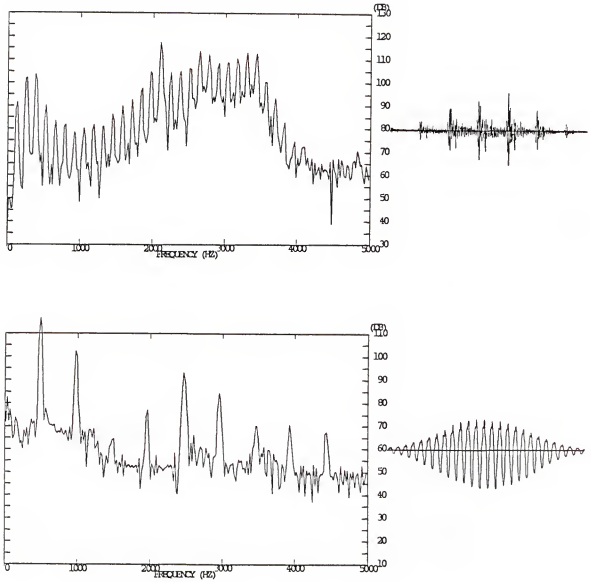


Figure 5-3 The speech spectra and the speech waveforms of a typical modal phonation (the upper graph) and a typical falsetto phonation (the lower graph). The phonating tasks were an /i/ by a male.

The pitch control mechanism in falsetto register is different from that in modal voice. Van den Berg [1968] states that:

> In modal voice the passive tension of the vocal ligaments needs to remain small when active tension in the vocalis muscles is increased to attain the highest pitches. In falsetto voice, however, the active tension in the vocalis muscles needs to remain small when the passive tension in the vocal ligaments is increased by the cricothyroid muscle to attain the highest pitches. The registers overlap in the region of medium pitches (pp. 293).

Hollien and Michel [1968] found that the average pitch range for male falsetto is 275-634 Hz, while the average range modal voice is 94-297 Hz.

The Vibratory Patterns of the Three Registers

The vibratory behaviors of the three registers are studied here based on two glottographic signals available: the EGG and the estimated volume velocity waveform. There is no glottal area measurement in our database for register study; instead, estimated glottal volume velocity from inverse-filtering [Lee, 1988] is used for determining instants of glottal opening and closing. Glottal volume velocity and glottal area may differ in skewness due to the source-tract interaction, but they should have similar open/closed phase divisions. Typical registrations for different vocal registers are shown in a

series of Figures (Figure 5-4 to 5-6). In each Figure, the graphs represent from top to bottom (1) the electroglottogram, (2) the estimated glottal volume velocity from inverse filtering, and (3) differentiated glottal volume velocity.

Figure 5-4 is the registration of a modal phonation (DMH). The glottal volume velocity waveform shows a distinct closed phase of about 40% of the entire pitch period. The large negative peaks in differentiated volume velocity indicate that the greatest excitation is located at the closing instant. The EGG waveform exhibits a steep falling slope and a gradual rising slope. The instant of maximum rising slope and the instant of maximum falling slope in the EGG waveform approximately match the opening and closing times, respectively, in the glottal volume velocity.

The registration of a typical falsetto phonation is shown in Figure 5-5. Compared to the modal phonation, the falsetto register does not have a distinct closed phase. This suggests that the glottis may remain open throughout the entire pitch period. Compared to the EGG waveform of a breathy phonation (Figure 5-7), the EGG waveform of a falsetto phonation has the shorter duty cycle, which translates into a more adductive glottis than in a breathy phonation. The finding that the glottis often remains slightly open in a falsetto phonation has led a number of researchers to suggest that a falsetto voice is usually accompanied by 'friction noises' or 'breathiness' [Zemlin, 1964]. In fact, given that the width

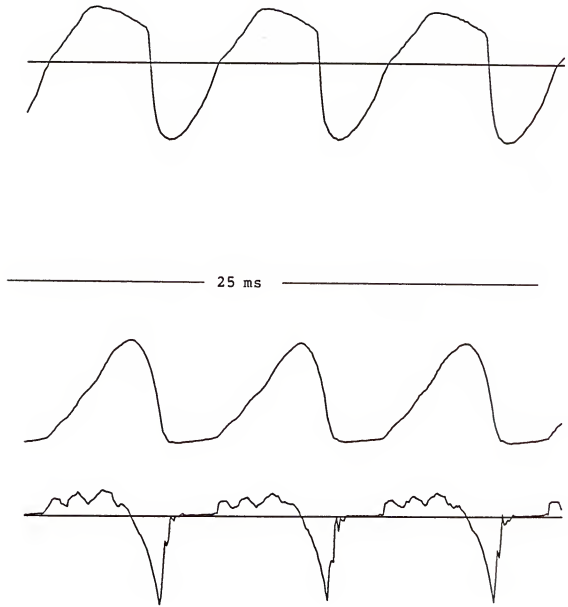


Figure 5-4 The glottographic waveforms of a typical modal phonation. These are, from top to bottom, the measured electroglottogram, the estimated glottal volume velocity, and the estimated differentiated glottal volume velocity. The phonating task was an /i/ by a male.

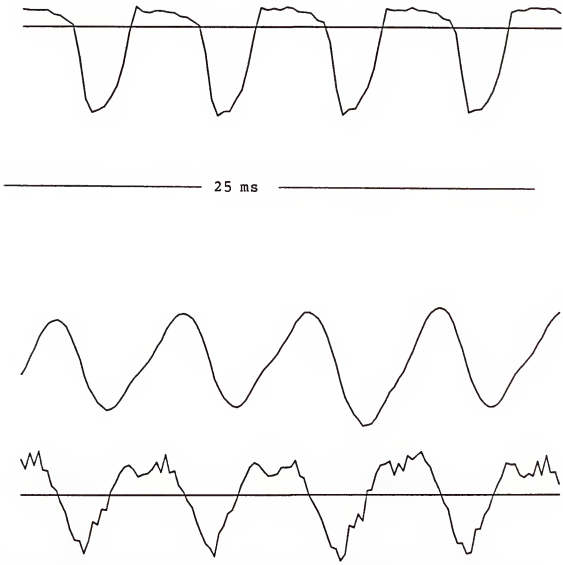


Figure 5-5 The glottographic waveforms of a typical falsetto phonation. These are, from top to bottom, the measured electroglottogram, the estimated glottal volume velocity, and the estimated differentiated glottal volume velocity. The phonating task was an /i/ by a male.

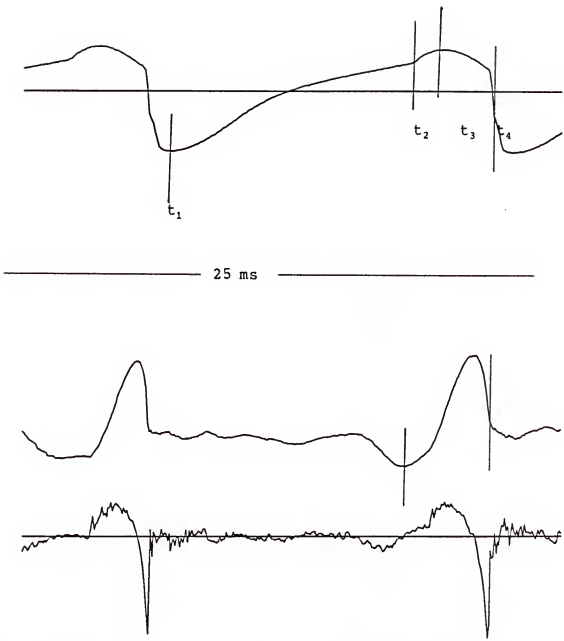


Figure 5-6 The glottographic waveforms of a typical vocal fry phonation. These are, from top to bottom, the measured electroglottogram, the estimated glottal volume velocity, and the estimated differentiated glottal volume velocity. The phonating task was an /i/ by a male.

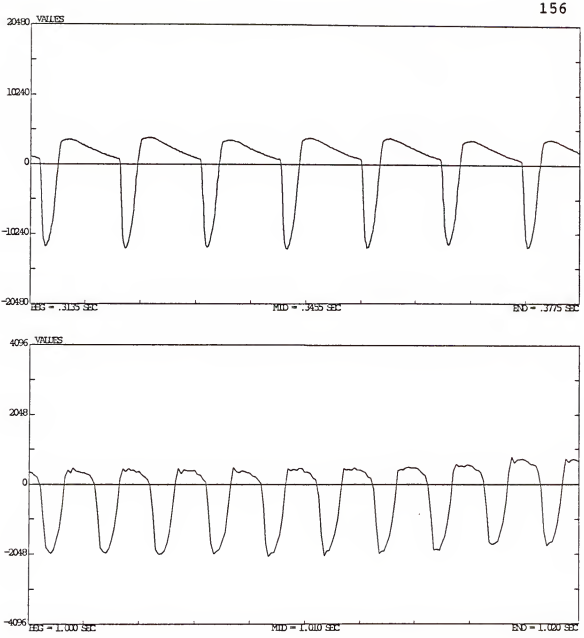


Figure 5-7 The electroglottographic waveforms of a breathy phonation (the upper graph) and a falsetto phonation (the lower graph). The phonating task was an /i/ by a male.

of opening is small, this fricative component is much more likely to be of the whispery type rather than the breathy type. On the other hand, the whispery effect may be slight if transglottal airflow is small, which is the case in a falsetto phonation. Lee [1988] found that the noise component in falsetto phonation is relatively small.

The absolute amplitude of the EGG signal is much smaller than that in a modal register, probably due to the thinner edges of the vocal folds and the fact that only part of the vocal folds is involved in vibration. Both the glottal volume velocity waveform and the EGG waveform are more symmetrical than those of a modal phonation.

Figure 5-6 is a typical registration of a vocal fry phonation. The most prominent feature is its short open phase (t_2 to t_4), which is about 15% of the entire pitch period. Both EGG waveform and glottal volume velocity exhibit steep falling slopes. The EGG waveform also has a very long and smooth rising slope (t_1 to t_3), but the glottis does not open until the time when an abrupt change of the slope occurs (t_2). Two phenomena may contribute to the long delay before opening. First, a large vertical phase difference exists between the upper edges and the lower edges of the vocal folds so that these edges alternately shade the glottis most of the time. Second, the folds are tightly pressed together for vocal fry when they collide; thus, the resulting area of contact starts

decreasing in the vertical dimension long before the folds really separate.

From the discussion above, it can be noted that the three vocal registers differ not only in range of fundamental frequency but also in vibratory patterns. These differences in vibratory patterns result in different spectral characteristics of the glottal source, and thus different vocal qualities.

Simulation of Glottographic Waveforms for the Three Vocal Registers

The purpose of this section is to provide a verification for the observed vibratory characteristics of the registers by reconstructing the corresponding glottographic waveforms and comparing the measured glottographic waveforms to the synthetic ones. The proposed vibratory model is set to simulate the vibratory behavior for the vocal registers for this purpose.

Modal Register

The vocal fold vibratory pattern of a modal phonation is characterized by a moderate fundamental frequency, wide lateral excursions, and a complete glottal closure. The asymmetry and duty cycle of the displacement function may vary according to pitch and vocal intensity.

For simulation of a typical modal phonation, the proposed model is set in the following conditions: CONF = 2510 (see Figure 3-6-b), DCAC = 0.6, PD = 50°, and HPD = 0°. The resulting glottal area (Figure 5-8) looks like a half sinusoid with an OQ of 0.55. The EGG waveform tends to skew to the right with a steep trailing slope, which coincides with the instant the glottal area closes.

Vocal Fry

Vocal fry register is characterized by tight adduction and a large phase difference between the upper and the lower edges of the glottis. Therefore, a longitudinal equilibrium configuration, shown in Figure 5-9, and a phase difference of 120 degrees were used for the simulation of vocal fry phonation. Shown in Figure 5-10 are the speech and the EGG waveform of a vocal fry phonation of vowel /i/ by a male subject. The dip on the top of the EGG waveform suggests the existence of a double-pulse vibration. Thus, the displacement function of the double-pulse type generated by the two-mass model in Chapter 2 is used here as the displacement function of the folds. The simulated glottal area function and the EGG waveform are shown in Figure 5-11. The OQ based on the area function is about 55% of the entire pitch period. This is much longer than that of the vocal fry phonation in the previous section (15%). This Figure is still reasonable, however, considering the time needed to produce double pulses.

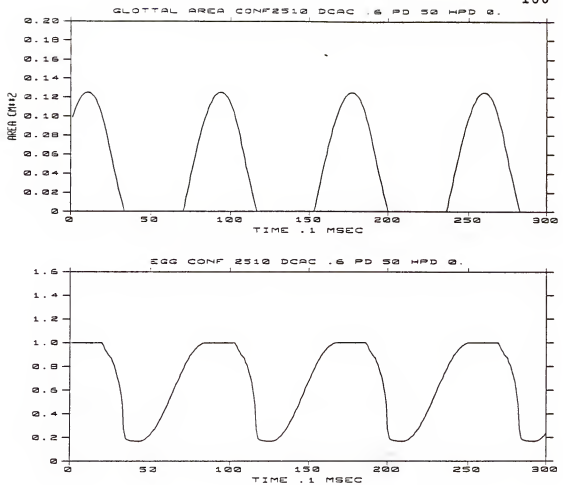


Figure 5-8 Simulated glottal area (the upper graph) and EGG waveform (the lower graph) for the following conditions: CONF =2510 (See the next figure), DCAC = 0.6, PD = 50, and HPD = 0.

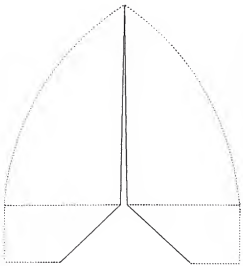


Figure 5-9 A longitudinal equilibrium configuration used for a simulation of vocal fry phonation.

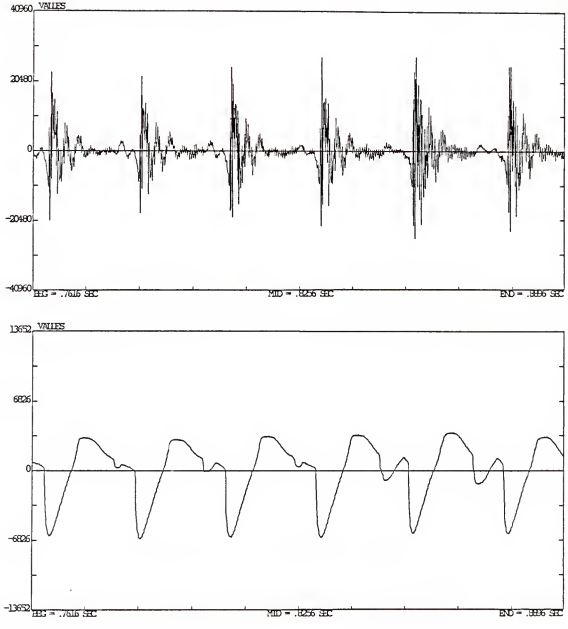


Figure 5-10 The speech (the upper graph) and the EGG waveforms (the lower graph) of a vocal fry phonation of vowel /i/ by a male subject.

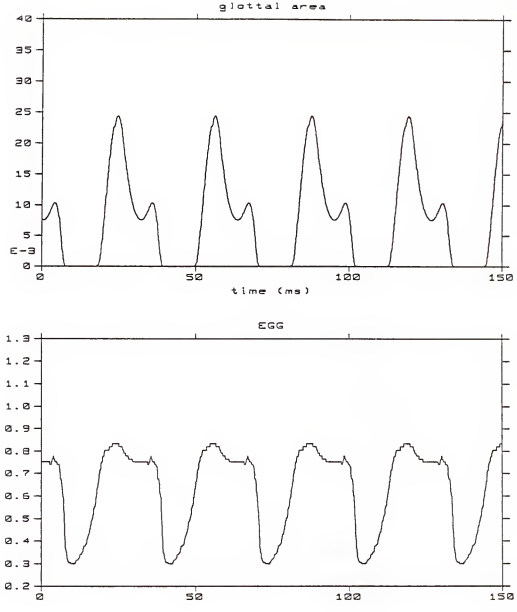


Figure 5-11 The synthetic glottal area function (the upper graph) and the EGG waveform (the lower graph) intended to simulate the vocal fry phonation shown in Figure 5-10. The phonating tasks were an /i/ by a male.

Falsetto

The OQ of the EGG waveform for falsetto phonation is usually larger than in modal phonation due to the relatively abducted vocal folds during falsetto. The EGG waveforms of a falsetto phonation and a modal phonation shown in Figure 5-12 provide an example for this trend. Also, according to simulation with the two-mass model, the displacement function of the masses (Figure 5-13) tends to skew to the left when the tension factor is high. We can see this tendency in the measured EGG waveform of a falsetto phonation (Figure 5-14).

Based on these factors, the simulation for a falsetto phonation is set up in the following conditions:

.a right-skewing displacement function; SQ = 0.7.

.wide abduction with small vibrating excursion; DCAC = 0.8.

.the longitudinal amplitude profile shown in Figure 5-15 in which only the middle portion of the folds is involved in the vibration.

.small phase difference; PD = 20°.

The simulated glottal area and EGG waveform, along with the measured EGG waveform of a falsetto phonation, are shown in Figure 5-16. The simulated EGG and measured EGG waveform match in most aspects except that the rising slope of the simulated EGG is slightly steeper than that of the measured EGG. The simulated glottal area waveform reveals complete closure does not occur and there is a skewing to the left.

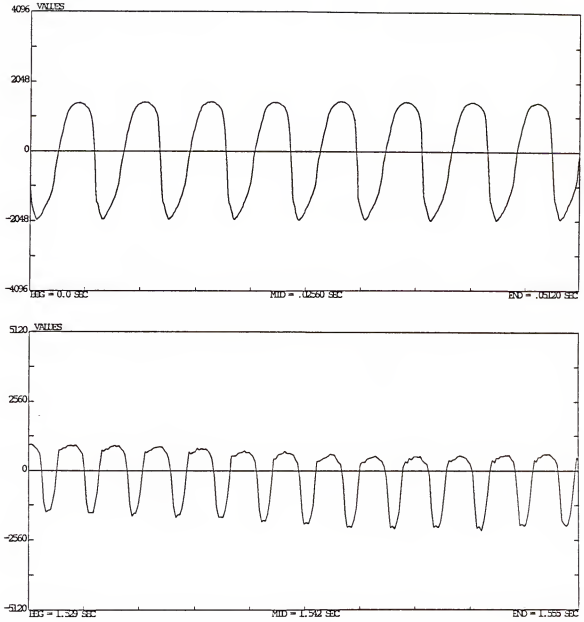


Figure 5-12 The EGG waveforms of a modal phonation (the upper graph) and a falsetto phonation (the lower graph). The phonating tasks were an /i/ by a male subject.

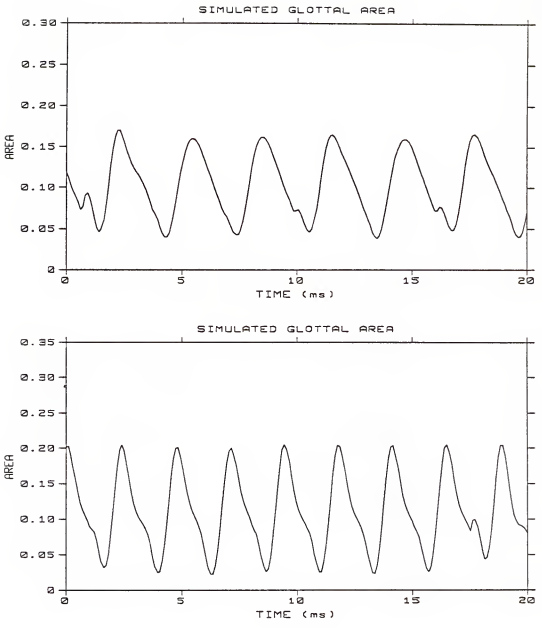


Figure 5-13 Simulated glottal area for tension factors 3 (for the upper graph) and 4 (for the lower graph).

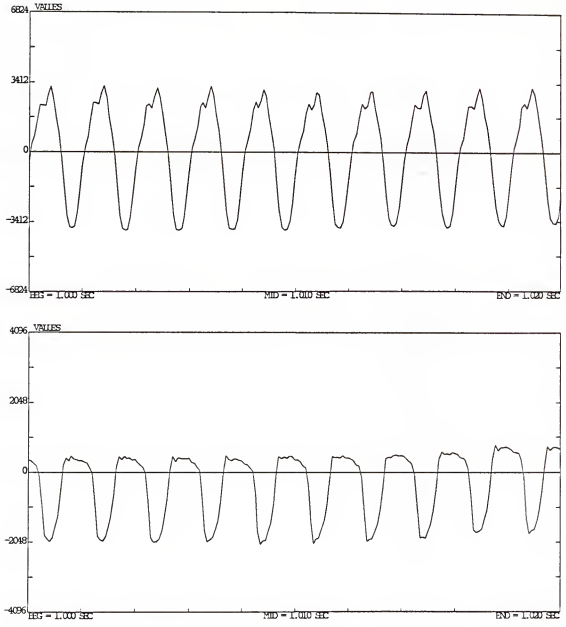


Figure 5-14 The EGG waveforms of two falsetto phonations. The phonating tasks were an /i/ by two male subjects.

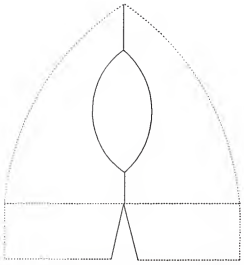


Figure 5-15 The longitudinal amplitude profile for a simulation of falsetto phonation.

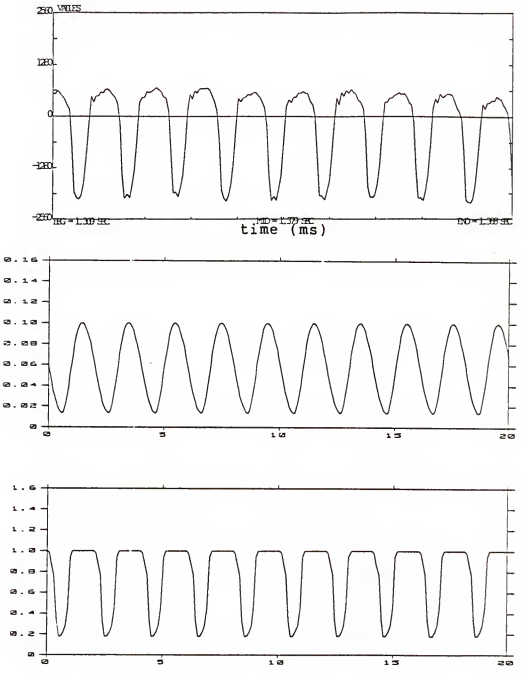


Figure 5-16 The measured EGG waveform of a falsetto phonation (upper), the simulated glottal area (middle), and the simulated EGG waveform (lower). The phonating task was an /i/ by a male.

Summary

In this chapter, we have shown that analysis and synthesis of voices of different registers involve more than a scaling of the fundamental frequency. Some basic differences in the phonatory and vibratory mechanisms were presented and discussed. Differences in glottal pulse shape, vocal tract damping, and glottal opening/closing gestures, together with the specific range of fundamental frequencies, separate one voice quality from another. The features of glottographic waveform for the three vocal registers are summarized in Table 5-1.

These glottographic waveform features of the three vocal registers were verified through a synthesis procedure for the glottographic waveforms by incorporating proper kinematic parameters into the proposed vibratory model.

Table 5-1. Summary of features of the glottographic waveforms for the three vocal registers

FEATURE	VOICE TYPE	MODAL VOICE	VOCAL FRY	FALSETTO
GLOTTAL AREA	OQ	medium	low	high
	skewing	no	right	left
	abruptness of closure	high	very high	low
EGG WAVEFORM	OQ	medium	low	high
	skewing	right	right	left
	irregular pattern	no	1/2 sub-harmonic	no

CHAPTER 6

CONCLUSIONS AND DISCUSSIONS

Summary

This research has developed a unique kinematic model that characterizes the vibratory motion of the vocal folds in terms of a three-dimensional glottal configuration with spatially varying tissue properties. We then used this model to: 1) establish a relation between the vibratory pattern of the vocal folds and the glottographic waveforms; 2) model the influence of voice disorders on the vibratory pattern of the vocal folds; and 3) study the behavior of glottographic waveforms associated with various vocal registers. The major achievements of this research are as follows.

The Kinematic Vibratory Model of the Vocal Folds

The proposed model is, as was Titze's vibratory model [1984], aimed at describing the kinematics of vocal fold vibration. Compared to Titze's model, the proposed model is more flexible in describing a 3-dimensional configuration of human vocal folds, incorporates more vibratory features, and allows simulations for irregular vibratory patterns and asymmetric fold properties. The flexibility of the vibratory

model allows simulations of the following situations: 1) the irregular boundary of the glottis, or protrusion due to a nodule, and/or polyp; 2) the changing maximum excursion along the length and depth of the glottis; 3) the anterior and the posterior portions of the vocal folds vibrate out of phase (longitudinal phasing); 4) the lower and the upper edges of the folds vibrate out of phase (vertical phasing); 5) the left and the right folds vibrate out of phase.

Relation Between the Vibratory Pattern and the Glottographic Waveforms

To interpret frequently used glottographic signals in relation to the vibratory pattern of the vocal folds, algorithms have been developed to express the three glottographic waveforms, electroglottogram, photoglottogram, and inverse-filtered glottal waveform, in terms of the displacement of the glottis. The simulations performed showed how the vibratory patterns of the vocal folds affect the waveshapes of glottographic waveforms. In particular, we found the asymmetry in the EGG waveform (skewing to the right) resulted from a combination of a convergent vertical shape of the glottis and the vertical phasing. We also found that the asymmetry of the glottal area waveform could be caused by different amplitudes of vibration for the superior and the inferior margins of the vocal folds as well as by vertical phase differences.

The source spectrum, which is closely related to the perception of voice quality, is greatly influenced by the vibratory pattern of the vocal folds. We demonstrated that the vibratory parameters, symmetry of displacement, abruptness of glottal closure, and abduction of the glottis, strongly influence the spectral characteristics of the vocal source.

Simulation of Abnormal Vocal Fold Vibration

Factors contributing to abnormal vibratory patterns, such as tension imbalance of the vocal folds, irregular shape of the vocal folds, and abnormal tissue properties, are investigated and associated glottographic waveforms are simulated with the proposed vibratory model. Two common vibratory patterns of abnormal vocal folds, i.e., incomplete glottal closure and perturbed vibration, were discussed and their underlying physiology and their influences on the characteristics of the corresponding glottographic waveforms were identified.

Since disease and organic changes may not affect both vocal folds in the same way, we investigated the influence of the following lateral asymmetries when they are altered in only one of the folds: unilateral nodule, tension imbalance, and unilateral paralysis. The result shows that an EGG waveform is more sensitive to irregular shape of vocal folds. Growths such as tumors or nodules can cause abrupt change in EGG waveform either in the opening phase or in the closing

phase. A tension difference between the two folds can cause an abrupt change in the falling slopes of the glottal area and the EGG waveform. Unilateral paralysis of the vocal folds can result in more gradual slopes at the opening and the closing in both glottographic waveforms.

Vibratory Pattern of Vocal Folds and the Resulting Vocal Quality

The basic differences in the phonatory and vibratory mechanisms among the vocal registers, that is, modal, vocal fry, and falsetto, were presented and discussed. Differences in glottal pulse shape, vocal tract damping, and glottal opening/closing gestures, together with the specific range of fundamental frequency distinguish one vocal register from another. These glottal waveform features of the vocal registers were verified through a synthesis of the glottographic waveforms by incorporating proper kinematic parameters into the proposed vibratory model.

Suggestions for Future Research

The proposed vibratory model of the vocal folds has demonstrated its ability to describe the complicated motion of human vocal folds. Nevertheless, there is still much room for improvement and extension. Further research is suggested in several areas.

1) The EGG, PGG, and inverse-filtered glottal waveforms, have been related to lateral contact area, glottal area, and glottal volume velocity, respectively. In this research, the proposed model has expressed these glottal characteristics in terms of the displacement of the glottis. To establish the relation between the glottographic waveforms and the vibratory pattern of the vocal folds, we still need research that determines the extent to which the EGG represents vocal fold contact area and the PGG represents projected glottal area.

2) Throughout this research, the performance of the proposed vibratory model has been judged by matching time-domain characteristics, (e.g., symmetry, OQ, slopes at discontinuous points, and overall waveshape), between the measured and the simulated glottographic waveforms. We are still unable to determine if these characteristics are of equal importance. For a specific application, some of the characteristics may be more critical than others. Therefore, there is a need to evaluate how well a vibratory model performs for a specific application. For voice synthesis, the source spectral characteristics should be of concern. For study of the role of the larynx in spoken language, the timing of glottal opening/closing gestures may be the focus.

3) Since the proposed vibratory model describes the motion of the vocal folds at the kinematic level and does not

explore the self-oscillatory nature of vocal fold vibration, a more comprehensive model with control parameters at the muscular level will be a logical follow-up of this research.

4) The proposed model has been applied to the study of the vibratory pattern of abnormal vocal folds and vocal registers. Also, subtle differences in laryngeal behavior are believed to be crucial in differentiating speakers of different sex and age. The proposed vibratory model should be able to contribute to such research.

5) The inclusion of a three-dimensional animation of the motion of the vocal folds offers: 1) more insights into the vibratory mechanisms because people tend to relate easily to graphic representation, 2) a visual feedback when the behavior of the vibratory model changes, and 3) a more natural way to make comparisons between the high-speed film data and the simulated vibratory motion.

REFERENCES

Abramson, A.S. (1979) "Laryngeal Timing in Consonant Distinctions," Phonetica 34, pp. 295-303

Ananthapadmanabha, T.V. (1984) "Acoustic Analysis of Voice Source Dynamics," Speech Transmission Laboratory, Quarterly Report 3, pp. 1-24

Ananthapadmanabha, T.V. and Fant, G. (1982) "Calculation of True Glottal Flow and Its Components," Speech Communication 1, pp. 167-184

Atal, B.S. and Remde, J.R. (1982) "A New Model of LPC Excitation for Producing Natural-Sounding Speech at Low Bit Rates," IEEE Int. Conf. Acoust., Speech and Signal Processing, pp. 614-617

Baer, T. (1975) "Investigation of Phonation Using Excised Larynxes," Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge.

Baer, T. (1979) "Reflex Activation of Laryngeal Muscles by Sudden Induced Subglottal Pressure Changes," J. Acoust. Soc. Am. 65, pp. 1271-1275

Berg, J. van den (1968) "Mechanism of the Larynx and the Laryngeal Vibrations," in Manual of Phonetics Edited by Malmberg, J. (1968), North-Holland, London, pp. 278-308

Berg, J. van den, and Tan, T.S. (1959) "Results of Experiments with Human Larynges," Oto-Rhino-Layrng. (Basel) 21, pp. 425-450

Berg, J. van den, Zantema, J.T., and Doornenbal, P. Jr. (1957) "On the Air Resistance and the Bernoulli Effect of the Human Larynx," J. Acoust. Soc. Am. 29, pp. 626-631

Boone, D.R. (1971) The Voice and Voice Therapy, Prentice-Hall, Englewood Cliffs, NJ.

Catford, J.C. (1964) "Phonation Types: the Classification of Some Laryngeal Components of Speech Production," Folia Phoniatrica 12, pp. 26-37

Catford, J.C. (1977) Fundamental Problems in Phonetics, Edinburgh University Press, Edinburgh

Childers, D.G., Hicks, D.M., Moore, G.P. and Alsaka, Y.A. (1986) "A Model for Vocal Fold Vibratory Motion, Contact area and the Electroglottogram," J. Acoust. Soc. Am. 80, pp. 1309-1320

Childers, D.G. and Krishnamurthy, A.K. (1985) "A Critical Review of Electroglottography," CRC Critical Review in Biomedical Engineering 12, pp. 131-161

Childers, D.G., Naik, J.M., Krishnamurthy, A.K., and Moore, G.P. (1983) "Electroglottography, Speech, and Ultra-high Speed Cinematography," Chapter 17 in Vocal fold Physiology and Biophysics of Voice, edited by I.R. Titze, Iowa City

Childers, D.G., Smith, A.M., and Moore, G.P. (1984) "Relationships Between Electroglottograph, Speech, and Vocal Fold Contact," Folia Phoniatrica 36, pp. 105-118

Coleman, R.F. (1963) "Decay Characteristics of Vocal Fry," Folia Phoniatrica 15, pp. 256-263

Damste, H., Hollien, H., Moore, G.P., and Murry, T. (1968) "An X-Ray Study of Vocal Tract Length," Folia Phoniatrica 25, pp. 349-359

Davis, S.B. (1979) "Acoustic Characteristics of Normal and Pathological Voices," in Speech and Language: Research and Theory, pp. 97-115, edited by N.K. Lass, New York, Academic Press

Fant, G. (1979a) "Glottal Source and Excitation Analysis" Speech Transmission Laboratory, Quarterly Report 1, pp. 85-109

Fant, G. (1979b) "Vocal Source Analysis - A Progress Report," Speech Transmission Laboratory, Quarterly Report 3, pp. 31-54

Fant, G. (1986) "Glottal Flow: Models and Interaction," J. Phonetics 14, pp. 393-399

Fant, G., and Liljencrants, J. (1985) "A Four-parameter Model of Glottal Flow," presented at the Vocal fold Physiology Symposium, New Haven, Conn.

Fink, B.R. and Demarest, R.J. (1978) Laryngeal Biomechanics Harvard University Press, Cambridge

Flanagan, J.L. (1958) "Some Properties of Glottal Sound Source," J. Speech Hear. Res. 1, pp. 99-116

Flanagan, J.L. (1968) "Self-Oscillating Source for Vocal-Tract Synthesizers," IEEE Trans. Audio and Electroacoustics, AU-10, pp. 57-64

Flanagan, J.L., Ishizaka, K., and Shipley, K.L. (1975) "Synthesis of Speech from a Dynamic Model of the Vocal Cords and Vocal Tract," Bell Syst. Tech. J. 54, (3), pp. 485-506

Flanagan, J.L., Rabiner, L.R., Christopher, D., Bock, D.E., and Shipp, T. (1976) "Digital Analysis of Laryngeal Control in Speech Production," J. Acoust. Soc. Am. 60, pp. 446-455

Fourcin, A.J., Wechsler, E., and Neil, W.F. (1976) "Laryngographic Analysis of Pathological Vocal Fold Vibration," Proceeding of the Institute of Acoustics, pp. 2-16-1 - 2-16-2

Fujimura, O. (1979) "Control of the Larynx in Speech," Phonetica 34, pp. 280-288

Gauffin, J., Sundberg, J., and Hammarberg, B. (1980) "Perceptual and Acoustic Correlates of Abnormal Voice Qualities," Octa Otolaryngol 90, pp. 441-451

Gauffin, J., Binh, N., Ananthapadmanabha, T., Fant, G. (1983) "Glottal Geometry and Volume Velocity Waveform," in Vocal Fold Physiology: Contemporary Research and Clinical Issues, edited by D. Bless and J. Abbs, New Haven, CT

Gray, G.W. and Wise, C.M. (1959) The Bases of Speech, New York, Harper

Hammarberg, B., Fritzell, B., Gauffin, J., and Sundberg, J. (1980) "Perceptual and Acoustic Correlates of Abnormal Voice Qualities," Acta Otolaryngolica 90, pp. 441-451

Hammarberg, B., Fritzell, B., Gauffin, J., and Sundberg, J. (1985) "Acoustic and Perceptual Analysis of Vocal Dysfunction," Presented at the Symposium on Voice Acoustics and Dysphonia, Gotland, Sweden.

Hiki, S., Imaizumi, S., and Hirano, M. (1976) "Acoustic Analysis for Voice Disorders," ICASSP, pp. 613-616

Hildebrand, B.H. (1976) "Vibratory Patterns of the Human Vocal Cords During Variations of Frequency and Intensity," Ph.D. dissertation, University of Florida, Gainesville

Hirano, M. and Koike, Y. (1974) "An Apparatus for Ultra-High-Speed Cinematography of the Vocal Cords," Ann. Otol. 83, pp. 12-18

Hirano, M. (1975) "Phonosurgery. Basic and Clinical Investigations," Otologia Fukuoka 21, pp. 249-440

Hirano, M., Matsuo, K., Kakita, Y., Kawasaki, H., and Kurita, S. (1983) "Vibratory Behavior vs. the Structure of the Vocal Fold," in Vocal Fold Physiology, Edited by I. R. Titze, Iowa City, pp. 26-40

Hirano, M., Vennard, W. and Ohara, J. (1970) "Regulation of Register, Pitch, and Intensity of Voice: An Electromyographic Investigation of Intrinsic Laryngeal Muscles," Folia Phoniatrica 22, pp. 1-20

Hollien, H. (1974) "On Vocal Registers," J. of Phonetics 2, pp. 125-143

Hollien, H., Coleman, R., and Moore, G.P. (1968) "Stroboscopic Laminagraphy of the Larynx During Phonation," Acta Otolaryngologica 65, pp. 209-215

Hollien, M. and Colton, R.H. (1969) "Four Laminagraphic Studies of Vocal Fold Thickness," Folia Phoniatrica 21, pp. 179-198

Hollien, M., Damste, H. and Murry, T. (1969) "Vocal Fold Length During Vocal Fry Phonation," Folia Phoniatrica 21, pp. 257-265

Hollien, H., and Michel, J.F. (1968) "Vocal Fry as a Phonational Register," Journal of Speech and Hearing Research 11, pp. 600-609

Hollien, H. and Wendahl, R.W. (1968) "Perceptual Study of Vocal Fry," J. Acoust. Soc. Am. 43, pp. 506-509

Ishizaka, K. and Flanagan, J.L. (1972) "Synthesis of Voiced Sounds From a Two-mass Model of the Vocal Cords," The Bell System Technical Journal 51, pp. 1233-68

Ishizaka, K. and Matsudaira, M. (1972) "Fluid Metalkinchanical Considerations of Vocal Cord Vibration," (SCRL Monograph No. 8). Santa Barbara Speech Communication Research Laboratory

Isshiki, N. and Ishizaka, K. (1976) "Computer Simulation of Pathological Vocal Cord Vibration," J. Acoust. Soc. Am. 60, pp. 1193-1198

Kasuya, H., Ogawa, S., Mashima, K., and Ebihara, S. (1986) "Normalized Noise Energy as an Acoustic Measure to Evaluate Pathologic Voice," J. Acoust. Soc. Am. 80, (5), pp. 1329-1334

Kitzing, P. (1983) "Simultaneous Photo- and Electroglottographic Measurements of Voice Strain," Chapter 18 of Vocal Fold Physiology, edited by I.R. Titze, pp. 221-229, Iowa City

Kitzing, P. (1985) "Stroboscopy - a Pertinent Laryngeal Examination," J. Otolaryg. 14, pp. 151-157

Kitzing, P. (1986) "Glottography: the Electrophysiological Investigation of Phonatory Biomechanics," Acta Oto-Rhino-Laryngol. Belg. 40, pp. 863-878

Koike, Y. and Markel, J.D. (1975) "Application of Inverse Filtering for Detecting Laryngeal Pathology," Annals of Otology, Phonology and Laryngology 84, pp. 117-124

Krishnamurthy, A.K., and Childers, D.G. (1986) "Two-Channel Speech Analysis," IEEE TRans. Acoust., Speech, Signal Processing 34, pp. 730-743

Ladefoged, P. (1975) A Course in Phonetics, Harcourt Brace Jovanovich, New York.

Laver, J. (1980) The Phonetic Description of Voice Quality, Cambridge University Press, Cambridge

Lecluse, F.L.E., Brocaar, M.P., and Verschuure, J. (1975) "The Electroglottograph and Its Relation to Glottal Activity," Folia Phoniat 27, pp. 215-224

Lee, C-K (1988) "Voice Quality: Analysis and Synthesis," Ph.D. Dissertation, University of Florida, Gainesville

Lofqvist, A. and Yoshioka, H. (1980) "Laryngeal Activity in Swedish Obstruent Clusters," J. Acoust. Soc. Am. 68, pp. 792-801

Monsen, R.B. and Engebretson, A.M. (1977) "Study of Variation in the Male and Female Glottal Wave," J. Acoust. Soc. Am. 33, pp. 179-186

Monsen, R.B., Engebretson, A.M. and Vemula, N.R. (1978) "Indirect Assessment of the Contribution of Subglottal Pressure and Vocal-Fold Tension to Changes of Fundamental Frequency in English," J. Acoust. Soc. Am. 64, pp. 65-80

Moore, G.P. (1971) Organic Voice Disorders, Prentice-Hall, Englewood Cliffs, NJ

Moore, G.P. (1976) "Observation on Laryngeal Disease, Laryngeal Behavior, and Voice," Annals of Otology, Rhinology, and Laryngology 85, pp. 553-567.

Moore, G.P. and von Leden, H. (1958) "Dynamic Variations of the Vibratory Pattern in the Normal Larynx," Folia Phoniat. 10, 1958, pp. 205-238

Moore, G.P., White, F.D., and von Leden, H. (1962) "Ultra-high Speed Photography in Laryngeal Physiology," J. Sp. Hear. Dis. 27, pp. 165-171

Muta, H. (1985) "Analysis of Hoarse Voices Using LPC Method," Fourth International Vocal Fold Physiology Conference, New Haven, CT

Romanes, G. (1978) Cunningham's Manual of Practical Anatomy, Vol.3, "Head and Neck and Brain," Oxford University Press, Oxford

Rosenberg, A.E. (1971) "Effect of Glottal Pulse Shape on the Quality of Natural Vowels," J. Acoust. Soc. Am. 49(2), pp. 583-590

Rosenberg A.E. (1981) "An Interactive Model for the Voice Source," presented at the vocal fold Physiology Symposium, Wisconsin, Madison

Rosenberg, M. (1983) "A New Inverse Filtering Technique for Deriving the Glottal Air Flow Waveform During Voicing," J. Acoust. Soc. Am. 53, pp. 1632-1645

Rothenberg, M. and Mahshie, J.J. (1988) "Monitoring Vocal Fold Abduction Through Vocal Fold Contact Area," J. of Speech and Hearing Res. 31, pp. 338-351

Schroeder, M.R. and Atal, B.S. (1985) "Code-Excited Linear Prediction (CELP): High-Quality Speech at Very Low Bit Rates," IEEE Int. Conf. Acoust., Speech and Signal Processing, pp. 937-940

Sundberg, J. and Gauffin, J. (1979) "Waveform and Spectrum of Glottal Voice Source," in Frontiers of Speech Communication edited by Lindblom and Ohman, Academic Press, London, pp. 301-322

Tanabe, M. (1975) "Analysis of High-Speed Motion Pictures of the Vocal Folds," Folia Phoniat. 27, pp. 77-87

Teaney, D. and Fourcin, A.J. (1980) "The Electrolaryngograph as a Clinical Tool for the Observation and Analysis of Vocal Fold Vibration," in The Voice Foundation, New York

Timcke, R., von Leden, H., and Moore, G.P. (1958) "Laryngeal Vibrations: Measurements of the Glottal Wave. Part I. The

Normal Vibratory Cycle," Archives of Otolaryngology 68, pp. 1-19

Titze, I.R. (1973) "The Human Vocal Cords: A Mathematical Model," Phonetica 28, pp. 129-170

Titze, I.R. (1976) "On the Mechanics of Vocal Fold Vibration," J. Acoust. Soc. Am. 60(6), pp. 1366-1380

Titze, I.R. (1982) "Synthesis of Sung Vowels Using a Time-Domain Approach," in Transcripts of the Eleventh Symposium: Care of the Professional Voice, (V. L. Lawrence Ed.), The Voice Foundation, New York, 90.

Titze, I.R. (1984) "Parameterization of the Glottal Area, Glottal Flow, and Vocal Fold Contact Area" J. Acoust. Soc. Am. 75(2), pp. 570-580

Titze, I.R. (1988a) "The Physics of Small-Amplitude Oscillation of the Vocal Folds," J. Acoust. Soc. Am. 83, pp. 1536-1552

Titze, I.R. (1988b) "A Hybrid Model of Vocal Fold Vibration," Draft Submitted to J. Acoust. Soc. Am.

Titze, I.R. (1988c) "A Four Parameter Model of the Glottis and Vocal Fold Contact Area," Submitted to Speech Communication

Titze, I.R. and Strong, W.J. (1975) "Normal Modes in Vocal Cord Tissues," J. Acoust. Soc. Am. 57(3), pp. 736-744

Titze, I.R. and Talkin, D. (1979a) "A Theoretical Study of the Effects of Various Laryngeal Configurations on the Acoustics of Phonation," J. Acoust. Soc. Am. 66, pp. 60-74

Titze, I.R. and Talkin, D. (1979b) "Simulation and Interpretation of Glottographic Waveforms," ASHA Reports, 11, pp. 48-53

Wendahl, R.W. (1963) "Laryngeal Analog Synthesis of Harsh Voice Quality," Folia Phoniatrica 15, pp. 607-618

Whitehead, R.L., Mets, D.E., and Whitehead, B.H. (1984) "Vibratory Patterns of the Vocal Folds During Pulse Register Phonation," J. Acoust. Soc. Am. 75, pp. 1293-1297

Wong, D.Y., Markel, J.D., and Gray, A.H., Jr. (1979) "Least Square Glottal Inverse Filtering from the Acoustic Speech Waveform," IEEE Trans. on ASSP-27, pp. 350-355

Zemlin, W.R. (1964) Speech and Hearing Science, Stipes, Champaign, IL

BIOGRAPHICAL SKETCH

Kwei Chan was born in Taiwan, Republic of China, on December 25, 1956. He received his Bachelor of Science degree in June 1979, and Master of Science degree in June 1981, all in the Department of Electronic Engineering at National Chiao-Tung University, Hsin-chu, Taiwan. Since August 1984 he has been with the Department of Electrical Engineering at the University of Florida, where his primary interests were digital signal processing and speech processing.

I certify that I read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Donald G. Childers, Chairman
Professor of Electrical Engineering

I certify that I read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Charles C. Formby
Associate Professor of Speech

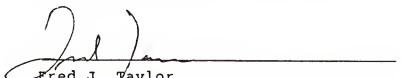
I certify that I read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Howard B. Rothman
Professor of Speech

I certify that I read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jack R. Smith
Professor of Electrical Engineering

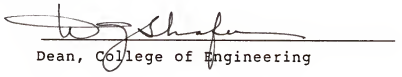
I certify that I read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Fred J. Taylor
Professor of Electrical Engineering

This dissertation was submitted to the Graduate Faculty of the College of Engineering and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1989



Dean, College of Engineering

Dean, Graduate School